~~1450~~
150
12.8.7

Psychological Bulletin

LYLE H. LANIER, Editor, *University of Illinois*
LORRAINE BOUTHILET, Managing Editor

4782 41

CONSULTING EDITORS

STEUART H. BRITT
Needham, Louis and Brorby, Inc., Chicago
DORWIN CARTWRIGHT
University of Michigan
FRANK A. GELDARD
University of Virginia
JAMES J. GIBSON
Cornell University
DAVID A. GRANT
University of Wisconsin
WILLIAM T. HERON
University of Minnesota
ERNEST R. HILGARD
Stanford University
WILLIAM A. HUNT
Northwestern University
JEAN WALKER MACFARLANE
University of California
DONALD G. MARQUIS
University of Michigan
JOHN T. METCALF
University of Vermont
JAMES G. MILLER
University of Chicago
NEAL E. MILLER
Yale University
HELEN PEAK
University of Michigan
ROBERT R. SEARS
Harvard University
ROBERT L. THORNDIKE
Teachers College, Columbia University

VOLUME 49, 1952

LIBRARY

PUBLISHED BIMONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH ST., N.W.,
WASHINGTON 5, D. C.

Edn. Research
12.8.70
J 150

CONTENTS OF VOLUME 49

ORIGINAL CONTRIBUTIONS, GENERAL REVIEWS, DISCUSSIONS

Thematic Apperception Test: Interpretive assumptions and related empirical evidence: GARDNER LINDZEY........... 1
When not to factor analyze: J. P. GUILFORD................. 26
A Note on Christie's: "Experimental naïveté and experiential naïveté": LEWIS BERNSTEIN.......................... 38
A note on the problem of homogeneity-heterogeneity in the use of the matching method in personality studies: PAUL F. SECORD.......................... 41
Tests of hypotheses: One-sided vs. two-sided alternatives: LYLE V. JONES.......................... 43
Some comments on Thistlethwaite's perception of latent learning: HOWARD H. KENDLER.......................... 47
The Blodgett and Haney types of latent learning experiment: Reply to Thistlethwaite: IRVING MALTZMAN.............. 52
Reply to Kendler and Maltzman: DONALD THISTLETHWAITE.... 61
Correction: A. RABIN and W. GUERTIN...................... 71
Item analysis in relation to educational and psychological testing: FREDERICK B. DAVIS.......................... 97
Non-parametric statistics for psychological research: LINCOLN E. MOSES.......................... 122
A computational short cut in factor analysis: R. V. ANDREE... 144
Note on "A qualification in the use of analysis of variance": C. H. PATTERSON.......................... 148
Comment on "A qualification in the use of analysis of variance": SOLOMON DIAMOND.......................... 151
A sequel to the notes of Patterson and Diamond: WILSE B. WEBB and VERNON LEMMON.......................... 155
The nature and measurement of meaning: CHARLES E. OSGOOD. 197
The validity of the Blacky Pictures: GERALD S. BLUM and HOWARD F. HUNT.......................... 238
A measure of relation determined by both mean difference and profile information: CHARLES E. OSGOOD and GEORGE J. SUCI.......................... 251
A note on functional relations obtained from group data: MURRAY SIDMAN.......................... 263
A note on the combination of ratings on the basis of reliability: C. H. LAWSHE and BRYANT F. NAGLE................ 270
The effects of psychological stress upon performance: RICHARD S. LAZARUS, JAMES DEESE, and SONIA F. OSLER........... 293

Influence of social class on personality test responses: FRANK AULD, JR. 318
The evaluation of experimentally controlled criteria: WILLIAM W. GRINGS 333
Reliable and questionable significance in a series of statistical tests: JOSEF BROŽEK and KENNETH TIEDE 339
The effectiveness of short methods for calculating correlation coefficients: JOHN C. FLANAGAN 342
Notice regarding change of editor 348
Publication Manual of the American Psychological Association. 389
Psychological tests in psychopathological prognosis: CHARLES WINDLE 451
Some observations on *Q* technique: WILLIAM STEPHENSON 483
The three basic factor-analytic research designs—their interrelations and derivatives: RAYMOND B. CATTELL 499
Some greco-latin analysis of variance designs for learning studies: E. JAMES ARCHER 521
A note on profile similarity: HAROLD WEBSTER 538
On the reliability of the leaderless group discussion technique: BORIS SEMEONOFF 540
Genetics of schizophrenia: A rejoinder: NICHOLAS PASTORE 542
The genetics of schizophrenia: Further rejoinder to Pastore: LEWIS A. HURST 544
A consideration of some conceptual trends in comparative psychology: T. C. SCHNEIRLA 559
The social behavior of vertebrates: A review of the literature (1939–1950): WENDELL I. SMITH and SHERMAN ROSS 598

BOOK REVIEWS

Blake and Ramsey's Perception: An approach to personality: DAVID C. MCCLELLAND 72
Ford and Beach's Patterns of sexual behavior: ELIOT STELLAR 75
Morgan and Stellar's Physiological psychology: JOSEPH E. BARMACK 78
Bartley's Beginning experimental psychology: CARL PFAFFMANN 80
Bugelski's A first course in experimental psychology: DONALD M. JOHNSON 82
Murphy's An introduction to psychology: FRANK W. FINGER 83
Garrett's Psychology: S. C. ERICKSEN 85
Pronko and Bowles' Empirical foundations of psychology: GEORGE F. J. LEHNER 86
Arnold's Das Raumlebnis in Naturwissenschaft und Erkenntnistheorie: KARL F. MUENZINGER 88
Cattell's Personality: A systematic, theoretical and factual study: JEAN WALKER MACFARLANE 89
Cattell's An introduction to personality study: ALFRED L. BALDWIN 90
Yacorzynski's Medical psychology. A basis for psychiatry and clinical psychology: SEYMOUR B. SARASON 92
Barker and Wright's One boy's day: A specimen record of behavior: CELIA BURNS STENDLER 92

Stevens' Handbook of experimental psychology: Lyle H. Lanier, David A. Grant, John L. Kennedy, John E. Anderson, Eliot Stellar, Judson S. Brown, Lorrin A. Riggs, W. R. Garner, and William E. Kappauf.... 156

Dollard and Miller's Personality and psychotherapy: James G. Miller and John M. Butler.... 183

Miller's Experiments in social process: A symposium on social psychology: Alfred F. Glixman.... 185

Thompson's Culture in crisis: Glen Heathers.... 187

Keller and Schoenfeld's Principles of psychology: A systematic text in the science of behavior: Judson S. Brown.... 189

Spearman and Jones's Human ability: D. R. Saunders.... 193

Lindquist's Educational measurement: Robert M. W. Travers.... 194

Tolman's Collected papers in psychology: Robert B. MacLeod.... 274

Katona's Psychological analysis of economic behavior: Samuel P. Hayes, Jr.... 276

Rohrer and Sherif's Social psychology at the crossroads: Harold Guetzkow.... 280

Queener's Introduction to social psychology: Clarence Leuba.... 282

Moreno's Sociometry, experimental method and the science of society: An approach to a new political orientation: John Harding.... 283

Volkart's Social behavior and personality. Contributions of W. I. Thomas to theory and social research: Kimball Young.... 284

Bohm's Lehrbuch der Rorschach-Psychodiagnostik für Psychologen Arzte und Pädagogen: Zygmunt Piotrowski.... 287

Helson's Theoretical foundations of psychology: Sigmund Koch.... 349

Marx's Psychological theory: William K. Estes.... 355

Munn's Psychology: The fundamentals of human adjustment: Stanley B. Williams.... 357

Stone's Comparative psychology: David Zeaman.... 359

Miller's Language and communication: C. E. Osgood.... 361

Luneburg's Mathematical analysis of binocular vision: P. Ratoosh and C. H. Graham.... 363

McClelland's Personality: Robert Leeper.... 367

Eells, Davis, Havighurst, Herrick, and Tyler's Intelligence and cultural differences: Quinn McNemar.... 370

Gurvitz's The dynamics of psychological testing: William A. Hunt.... 372

Janis' Air war and psychological stress: Psychological studies of bombing and civilian defense: M. Brewster Smith.... 372

Rogers' Client-centered therapy; its current practice, implications, and theory: Dwight W. Miles.... 374

Welford's Skill and age. An experimental approach: Nathan W. Shock.... 376

Berrien's Comments and cases on human relations: S. Stansfeld Sargent.... 377

Ellis' Educational psychology: Donald Snygg.... 379

English's Child psychology: T. W. Richards.... 380

Horrocks' The psychology of adolescence: Glenn M. Blair.... 381

Symonds' The ego and the self: Arthur Jenness.... 382

Culbertson's Consciousness and behavior: F. J. McGuigan.... 384

Symposia of the Society for Experimental Biology: IV. Physiological mechanisms in animal behavior: Frank A. Beach.... 385

Humphrey's Thinking: An introduction to its experimental psychology: George A. Miller.... 547

Kempthorne's The design and analysis of experiments: C. J. Burke.... 548

Jahoda, Deutsch, and Cook's Research methods in social relations: With especial reference to prejudice: J. C. GILCHRIST 549
Tilton's An educational psychology of learning: J. B. STROUD 551
Hull's Essentials of behavior: W. N. SCHOENFELD and P. J. BERSH 628
Parsons and Shils's Toward a general theory of action: SIGMUND KOCH 636
Tinbergen's The study of instinct: L. I. O'KELLY 642
Cameron and Magaret's Behavior pathology: SEYMOUR B. SARASON 644
Schneiders' Introductory psychology: The principles of human adjustment: SISTER M. AMATORA 645
Hofstätter's Die Psychologie und das Leben: GEORGE F. J. LEHNER 646
Dewey and Humber's The development of human behavior: DALE B. HARRIS 647
Kuhlen's The psychology of adolescent development: BOYD MCCANDLESS 649
Watson's The clinical method in psychology: GERALD S. BLUM 651
Pascal and Suttell's The Bender-Gestalt Test: Quantification and validity for adults: ELEANORE OCHS 652
Rapaport's Organization and pathology of thought: ROBERT P. FISCHER 653
Brunswik's The conceptual framework of psychology: GUSTAV BERGMANN 654
Stone and Taylor's Annual review of psychology: MELVIN H. MARX 657
MCGEOCH and IRION's The psychology of human learning: GREGORY A. KIMBLE. 660
Kelly and Fiske's The prediction of performance in clinical psychology: J. R. WITTENBORN 662

FILM REVIEWS

Pronko and Snyder's Vision with spatial inversion: DOROTHEA E. JOHANNSEN 553
Miller and Hunt's Motivation and reward in learning: TRACY S. KENDLER 554
Hayes and Hayes' Vocalization and speech in chimpanzees: EDWARD M. BENNETT 555
Stone's When should grown-ups help? and And then ice cream (children's meals): ELIAS KATZ 556
Verplanck's Testing intelligence with the Stanford-Binet: ELIAS KATZ 557
Bennett and McKeever's "Antabuse" in the treatment of alcoholism: ALBERT D. ULLMAN 557

MISCELLANEOUS

Books and Materials Received: 94, 289, 386, 665
Index of Subjects: 668
Index of Authors: 670

Vol. 49, No. 1 January, 1952

Psychological Bulletin

THEMATIC APPERCEPTION TEST: INTERPRETIVE ASSUMPTIONS AND RELATED EMPIRICAL EVIDENCE[1]

478 Z 41.

GARDNER LINDZEY
Harvard University

The chief purpose of this article is to state the assumptions customarily involved in interpreting the Thematic Apperception Test,[2] and to examine the logical considerations and some of the empirical evidence that can be used to verify or reject each of these assumptions.

Aside from certain historical ties to psychoanalysis, the theoretical and empirical continuity between projective testing and the remainder of psychology has been a subject of little interest. This is true in spite of an increasingly large amount of research and formulation in other areas of psychology that is directly pertinent to the activities involved in projective testing. In particular, the research of the "new-look" perceptionists represents an important and fertile link between projective testing and more traditional domains of psychology. The initial outline of such a continuity has been traced by Blake and Wilson (7) in a study of the influence of depressive tendencies upon selectivity in Rorschach response, Bruner (9) in a discussion of the Rorschach test, Lawrence (39) in an investigation of temporal factors in perception, Siipola (75) in an examination of the effect of color upon Rorschach response and Stein (76) in an ingenious study employing tachistoscopic exposure of Rorschach stimulus material. All of thse studies illustrate the feasibility and fruitfulness of relating perception findings to material elicited by Rorschach-like techniques. One of the aims of the present paper is to stress the desirability of relating such theory and research to the findings of investigators who employ the Thematic Apperception Test.

---

[1] This paper is an outgrowth of a study of personality and the imaginative processes directed by Henry A. Murray and supported by grants from Rockefeller Foundation and the Laboratory of Social Relations, Harvard University. I am deeply grateful to a number of colleagues for their generosity in reading and criticizing this manuscript. My greatest single debt is to Henry A. Murray whose constant stimulation and encouragement made this article possible.

[2] These same assumptions are customarily employed in connection with other story-construction projective techniques, e.g., Make-A-Picture-Story Test, Four-Picture Test, Tri-Dimensional Apperception Test, etc.

One may legitimately object to my proposal to appraise empirically statements that I am treating as given, or axiomatic. However, it is not clear that all psychologists would concur in giving axiomatic status to these statements. As is true of so much of psychological formulation, the distinction here between analytic and empirical is not clearly delineated—what is one person's empirical generalization is another's axiom. Further, I believe most psychologists agree with MacCorquodale and Meehl (**44**) in subscribing to a methodological position that emphasizes the use of *only* analytic constructs that interact smoothly with available empirical knowledge. Consequently, examination of these assumptions in the light of empirical evidence seems justified on the one hand by the questionable axiomatic status of the assumptions and on the other by the general acceptance of psychologists that axioms should not violate observational data.

The assumptions to follow vary greatly in their generality. In fact, they are logically related only by virtue of their frequent use in the interpretation of material secured from one particular type of projective technique. They do not represent all of the assumptions employed in making such interpretations. I have attempted to include only those that are not sufficiently general to be common among all psychologists and yet, within the prescribed situation, are sufficiently common so that almost any individual engaged in this activity would employ them. For the latter reason, I am not concerned here with the assumptions involved in "formal" or "sign" analysis of story projective material where the meaning, or the thematic qualities, of the material is minimized, e.g., Balken and Masserman (**3**) and Wyatt (**84**). Nor have I attempted to explore the assumptions involved in the more recent "sequence analysis" (**1**) of projective responses.

THE ASSUMPTIONS AND RELATED EVIDENCE

The assumptions are divided into three crude groups. First, the most general assumption, fundamental to all projective testing. Second, those assumptions that are concerned with procedures employed in determining the diagnostically significant portions of the fantasy productions. Third, those assumptions involved in relating the significant portions of the protocols to other forms of behavior.

Primary Assumption: In completing or structuring an incomplete or unstructured situation, the individual may reveal his own strivings, dispositions, and conflicts.[3]

---

[3] The terms "strivings, dispositions, and conflicts" are meant to designate all the attributes or aspects of the person that the clinician is interested in or wishes to measure. One could readily add to this list such terms as: personality organization, primitive fixations, complexes, needs and press, or any others that seemed necessary to represent those aspects of the person that are being explored.

Assumptions Involved in Determining Revealing Portions of Stories:

1. In the process of creating a story the story-teller ordinarily identifies with one person[4] in the drama, and the wishes, strivings, and conflicts of this imaginary person may reflect those of the story-teller.

   a. It is assumed further that the identification figure can be established through the application of a number of specific criteria; e.g., person appearing first in the story, person doing most of the behaving, person most similar to story-teller, etc.

   b. It is also assumed that additional figures in the stories such as father, mother, or brother often may be equated to the real-life counterparts of the story-teller and the behavior of the hero toward them used as indicative of the story-teller's reactions to these persons.

2. The story-teller's dispositions, strivings, and conflicts are sometimes represented indirectly or symbolically.

3. All of the stories that the subject creates are not of equal importance as diagnostic of his impulses and conflicts. Certain crucial stories may provide a very large amount of valid diagnostic material while others may supply little or none.

4. Themes or story-elements that appear to have arisen directly out of the stimulus material are less apt to be significant than those that do not appear to have been directly determined by the stimulus material.

5. Themes that are recurrent in a series of stories are particularly apt to mirror the impulses and conflicts of the story-teller.

Assumptions Involved in Deriving From Revealing Portions of Fantasy Material Inferences About Other Aspects of Behavior:

1. The stories may reflect not only the enduring dispositions and conflicts of the subject, but also conflicts and impulses that are momentarily aroused by some force in the immediate present.

   a. The further assumption is frequently made that both the enduring and temporary processes are reflected in stories in the same manner.

2. The stories may reflect events from the past of the subject that he has not himself actively experienced, but rather has witnessed or observed, e.g., street scene, story, motion picture.

   a. It is assumed further that, although the subject has not himself experienced these events, and is telling them as he observed them, the fact that he selects these events, rather than others, is in itself indicative of his own impulses and conflicts.

3. The stories may reflect group-membership or socio-cultural determinants in addition to individual or personal determinants.

4. The dispositions and conflicts that may be inferred from the story-teller's creations are not always reflected directly in overt behavior or consciousness.

I wish to emphasize that the research to be discussed in connection with the various assumptions can *not* serve as a validation of this kind of projective testing. At most, it can demonstrate that the very general assumptions lying behind this kind of activity are not in direct conflict with available empirical findings. This may increase the "plausibility" of projective interpretations, but the task of demonstrating the utility of specific rules of interpretation remains.

---

[4] Most investigators accept the possibility of multiple identifications. However, this necessitates the same assumption and only complicates somewhat the general problems of establishing interpretive rules.

It is likewise important to realize that in this article I am omitting from consideration most of the pertinent empirical evidence resulting from clinical use of this technique. The omission of this idiographic research or observation is not the consequence of any feeling that such data are not of immense value. Rather, in an article emphasizing the relation of projective testing to the remainder of psychology, it seemed desirable to stress most heavily those kinds of research that came closest to meeting the experimenter's demand for empirical control and intersubjectivity of method. Thus, most of the evidence to be referred to has been secured with some attempt at maintaining adequate empirical controls.

Does the Individual Reveal His Own Dispositions and Conflicts in Completing an Unstructured Situation?

This assumption is not limited to story-construction tests but lies at the heart of all projective testing. Fortunately, in view of its ubiquity, there is a host of very general experimental verification, some of it drawn from laboratory research.

One of the first investigators to point to the relationship between motivational factors and response in an unstructured situation was Sanford (**71**, **72**). He demonstrated in a series of experiments that the food responses of subjects varied as a function of the amount of food deprivation they had undergone. The responses were given in the act of telling stories to ambiguous pictures, making word associations, and in other situations where the stimulus was sufficiently unstructured to permit either food or non-food responses. This same general function was later demonstrated under somewhat different conditions by Levine, Chein, and Murphy (**40**) and McClelland and Atkinson (**48**). The results of these three investigations do not agree in detail, but all were able to demonstrate some kind of variation in response to ambiguous stimulus situations as a function of food deprivation.

In an early study by Murray (**55**), it was shown that estimates of "maliciousness" of faces in photographs varied directly with an experimentally induced state of fear. After a fear-producing game children rated a series of photographed faces as significantly more malicious than they had rated the faces previous to the game. A recent study by Katz (**33**) has shown that the manner in which an individual completes an incompletely drawn face varies systematically with the kind of experience he has undergone previous to the activity. Thus, a group that had just failed on a test of reasoning ability differed significantly from a group that had just spent their time in a neutral activity. Presumably the manner in which the individual completed this unstructured situation was directly related to the motivational or emotional conditions aroused by the experimental treatment.

A series of investigations by Bruner and Postman (12) have demonstrated that when stimulus material is made ambiguous by very brief tachistoscopic exposure, the responses or "guesses" of the subjects vary systematically with motivational variables. These same investigators, as a result of their perception research, have come to occupy a much more extreme position than that implied by the above assumption. They suggest that:

> Most experimenters who have worked with need and attitude factors in perception have assumed, sometimes quite explicitly, that only in highly equivocal stimulus situations can such "nonsensory" factors operate. . . . But all stimulus situations are potentially equivocal and cease to be so only to the extent that selection, accentuation, and fixation have taken place. Perception occurring without the contribution of such adaptive factors is as unthinkable as perception without the mediation of receptive nerve tissue (11, p. 301). . . . *Adaptive factors in perception are not limited to unstable stimulus situations* (11, p. 307).

In the light of this evidence, the assumption that motivational factors are revealed in completing unstructured situations seems clearly warranted.

Does the Subject Identify with Some Figure in the Stories He Tells and Can This Figure Be Established Reliably?

There seem to be three modal positions that can be maintained in regard to identification in the story-telling process. First, we may make the assumption, indicated above, that there is *ordinarily* a single identification figure in each story. This assumption can be complicated greatly by a number of special conditions of the type already suggested by Murray (56). Second, we may assume a continuum of identification with those figures in the stories that are very similar to the story-teller possessing a maximum of the subject's attributes and those that are very dissimilar possessing a minimum. This is the assumption made by Sears (74), who, working primarily with doll techniques and heavily influenced by the notion of "stimulus generalization," has suggested specifiable dimensions along which the degree of identification can be expected to vary. Thus, in doll play figures resembling the subject in age and sex will show more characteristics of the subject than those figures that are dissimilar in age and sex. Third, we might, with Henry (30), Piotrowski (58), and others, make no attempt to locate the hero but simply look on all characters in the constructed stories as representative of aspects of the story-teller. This last alternative is perhaps the least happy, not only because a very large amount of diverse clinical experience militates against it, but also because it leads to certain drastic limitations upon the diagnostic use of the TAT. If we adopt this assumption, we are more or less forced to give up the attempt to appraise the subject's attitudes toward other persons. Thus, Piotrowski suggests that we eliminate the hero and assume "that every figure in the TAT stories expresses some aspect of the testee's personality" (**58**, p. 107), while somewhat later he suggests that many stories "reflect what the subject thinks and feels about persons represented by the TAT figures, i.e., about the old and the young, the male and the female" (**58**, p. 113). Thus, with no rules for differentiating, we are told that all figures represent characteristics of the hero, but also that the characteristics of some figures represent the story-teller's attitudes toward other persons. If we assume that all figures are equally representative of the story-teller's characteristics, either we must give up attempts to appraise the subject's attitudes toward other persons through this instrument or else engage in some kind of dialectic in the effort to defend such attempts.

There is, as yet, no unequivocal answer to the question of which of the above assumptions is most useful. In order to show that the assumption of a single identification figure is warranted, it is necessary to demonstrate relationships between the behavior of the subject and the imaginary behavior of the hero that do not exist between the behavior of the subject and the behavior of non-hero figures. If this difference is not shown, one can always suggest that the sensitivity of the test to aspects of the subject is a result not of any specific identification process, but rather of a generalized reflection of motivational state of the sort that Murray (**55**), Bruner and Postman (**12**), and others have reported. Thus, even if *all* figures in the stories represented aspects of the story-teller, we might arbitrarily decide to use only a certain percentage of these figures as diagnostic of the story-teller's attributes, i.e., employ the hero and non-hero distinction, and we would still expect the test to show some sensitivity to variations in story-teller behavior even though we were wasting much of its power. Just as the subject's motivational states are reflected in free association, or in the pre-solution hypotheses of tachistoscopic response, so also the verbal flow accompanying the story-telling process may mirror motivational states without any intervening identification with particular actors in the stories.

The distinction between "heroes" or identification figures on the one hand and "non-heroes" on the other is greatly complicated by the fact that the story-teller's orientation or attitudes toward other persons is intimately related to his own psychic makeup. Thus, we might mistakenly employ an identification figure as representative of the story-teller's attitudes toward other people and actually secure considerable information about the individual's external orientation because of the similarity between "self" and "perceived other." Or, conversely, we might take a figure intended by the story-teller as an "other" object and find mirrored in it much of real pertinence to the story-teller's own personality structure.

What is needed to answer the above question is a clear demonstration of more intimate relations between "hero" behavior or attributes and the story-teller than can be shown between "non-hero" behavior and the story-teller. Results such as Bellak's (**5**) showing that the frequency of aggressive words rises when the individual is frustrated do not demonstrate that the identification figure is necessarily displaying more aggressive behavior. Nor are studies satisfactory that simply demonstrate variations in "hero" activities or attributes that relate to behavior or attributes of the story-teller. Even if Bellak had shown that aggressive behavior on the part of heroes rose following frustration, this would still not provide the needed information. In this case it is quite possible that analysis of the behavior of the non-identification figures would reveal the same increase in aggressive behavior, indicating not identification, but simply an increase in verbal responses pertinent to the need or conflict in question.

Lindzey (**43**) has demonstrated that following a social frustration the incidence of aggressive acts in TAT stories carried out by heroes against others increased more than the incidence of aggressive acts carried out by "other" figures. If the incidence of the two kinds of acts is combined, the resulting shift is greater than the shift in either "self" or "other" figures alone. These findings are complicated by the fact that the frustration situation was of such a nature that both extrapunitive responses and a view of the environment as hostile and threatening would be expected to rise following the experimental treatment. Thus, the increase in aggressive responses on the part of non-hero figures might be considered a result of the increase in aggressive tendencies on the part of this story-teller *or* a reflection of the fact that the story-teller viewed the external world as a more hostile and threatening place. It is clear that what is needed is an experimental treatment where, given the subject's identification with a hero-figure, the predictions to be made for "hero" and "other" figures will either be opposed or widely different.

Accepting the desirability of locating identification figures, can these figures be specified precisely? Several studies (**29**, **79**) in which relatively high inter-scorer reliability coefficients were secured imply that it *is* possible to obtain reliable agreement among different scorers in establishing the identification figure. Presumably, a positive correlation for need and press ratings could be secured only if the different raters were treating the same figures as hero. In addition, Mayman and Kutner (**45**) report complete agreement in 89 per cent of the cases between two raters who independently determined the identification figure for a series of 91 stories.

In general, then, the feasibility of the identification assumption cannot be clearly demonstrated at present, although empirical evidence suggests that identification figures can be established with reasonable reliability.

Are Impulses and Conflicts of the Subject Sometimes Represented Symbolically?

The clinical use of the concept of symbolism is intimately related to a number of psychological concepts employed by more rigorous and empirically oriented investigators. All behavior theorists have encountered the thorny problems posed by interchangeability or substitutability in behavior of stimuli and responses. These problems have led to the formulation of a number of psychological concepts including displacement, substitution, stimulus generalization, and vicarious mediation. Apparently most psychologists agree that when a response is interfered with either by internal or external barriers it will frequently become altered to another form or else be directed toward a new object. There is a large body of research, much of it carried out on animal subjects, demonstrating this relationship and exploring some of the conditions under which it operates (**8, 15, 25, 31, 32, 38, 62, 63, 83**). Of the animal investigations, perhaps the most pertinent to our discussion is Miller's (**54**) demonstration that a learned aggressive response can be generalized or displaced from an original stimulus object (another rat) to a substitute stimulus (doll) when the first stimulus is made unavailable. This process can be considered a rough paradigm of symbolic representation.

A special question of considerable importance to the projective tester is whether symbolic transformations can, and customarily do, take place without awareness of the subject. Most of the clinical studies discussed below deal with symbolic representations which the subject was not aware of initially and which he could become aware of only under special conditions, e.g., interpretation or therapeutic change. This capacity of the organism to engage in symbolic representation without awareness of the process has been demonstrated under better controlled circumstances by other investigators. Diven (**17**) and Haggard (**26**) have both shown that a word may become a substitute stimulus for an electric shock and evoke a galvanic skin response that differentiates it from "non-shock" words, even where the subject is unaware of the connection between the word and the electric shock. In similar fashion, McCleary and Lazarus (**46**) showed that words to which a galvanic skin response had been conditioned through the use of electric shock could elicit the galvanic skin response even when exposed tachistoscopically at such rapid speeds as to prevent conscious recognition.

Certainly since Freud, the clinical interpretation of diagnostic material, whether projective or not, has depended heavily upon the assumption of symbolic transformations. There is some clinical-experimental and a host of clinical-observational evidence to indicate that symbolic transformations are quite customary in human behavior, particularly where unacceptable or antisocial impulses are involved.

In case histories conversion symptoms and compulsions frequently present themselves dramatically as evidence of the fact that a given conflict or forbidden impulse may secure expression in a manner only indirectly related to the original impulse. Thus, it is generally accepted that hand-washing compulsions often represent symbolically the desire of the individual to cleanse himself of the consequences of masturbation. Likewise, in doll therapy there is excellent evidence that the child is able to discharge against surrogate objects the same impulses and feelings that he has developed toward persons in the real world. The extent to which this relationship between symbol and real-life counterpart is unequivocally demonstrated varies with the individual case. However, in many instances the relationship is effectively demonstrated as the symbolic behavior can be shown to vary directly with changes in the relationship between the subject and the referent of the symbol. Thus, the aggression directed against the father-doll may show a high inverse relationship to the extent to which the child is able to express such impulses against the original stimulus object—the father. Similarly, a symptom may disappear when the impulse or conflict that it is presumed to represent is dissolved by therapeutic procedure or through changes in the balance of environmental determinants. Levy (**42**) has described a number of clinical cases involving relatively convincing demonstration of symbolic representation of sibling rivalry, fear of castration, defecation, etc. Tomkins (**80**) reports the study of a single case under varying degrees of drunkenness. He found that in certain cases the hypothesized referent of the symbol was more and more directly represented as the amount of alcohol consumed increased.

Several investigators have attempted to combine the subject matter of clinical psychology with attempts at empirical control in the study of symbolic representation. An interesting study by Farber and Fisher (**21**) attempted to demonstrate the process of symbolic transformation by means of using hypnosis and asking subjects to create and interpret dreams. Although the study is not well controlled, if we accept the naïveté of the subjects, the evidence is impressive, for some of the subjects, that not only is symbolic representation an observed form of behavior, but also that the process of assigning symbols in large part corresponds with the rules established by Freud from his examination of dream protocols. Consistently, Krout (**37**) in a better-controlled investigation studied interpretations based upon the assumption that individuals responded to line drawings, representative of basic experiences or objects, *as if* they were the symbol referent. She found that these interpretations could be verified through the inspection of independently collected validation data. In other words, the subjects did appear to respond to the line drawings in a manner consistent with the nature of the object that the Freudian would view as lying behind this symbol. Franck (**22**) and Franck and Rosen (**23**) have reported two studies in which there is some evidence that the response of the individual to male and female symbols varies with his sex and relative maturity. Klein (**34**), employing the method of hypnotically inducing dreams, was able to show that the dreamer characteristically transformed or disguised external stimuli that were incorporated into the dream.

In general, then, controlled empirical evidence supplements clinical observation to imply the existence of tendencies for human subjects to represent dispositions and conflicts symbolically. The rules for determining the symbol-referent relations are as yet imperfectly understood, although there is evidence indicating the Freudian view of the symbolic process possesses some utility.

Are the Stories That the Subject Creates of Unequal Importance as Diagnostic of His Impulses and Conflicts?

It is possible to assume that any behavioral information, no matter how scanty, contains in it the necessary elements to permit a complete understanding of the individual in question. This derivative of strict Freudian determinism is still defended by some investigators. It implies a theoretical model that is able to incorporate all possible empirical relationships and predict these relationships so precisely that once a single value has been inserted into the closed theoretical system, every other construct in the system can be assigned a value. In view of the present state of psychological theory, this assumption seems an exceedingly unwise one. Nor does it appear to guide the actual behavior of most clinicians. The tendency to view different stories as possessing different degrees of psychological significance is rather clearly revealed by the fact that not infrequently, even with twenty stories, the investigator will feel he has not sufficient material to make a diagnosis or come to any understanding of the dynamics of the individual in question. Apparently just as some sets of twenty stories are more difficult to interpret and are less revealing, so also some individual stories are less rewarding than others.

Although presumably implicitly present in the minds of most interpreters, the assumption is seldom mentioned and there has been little systematic attempt to state any criteria by means of which the more important stories are to be separated from the less important. Rapaport (**61**), who represents an exception to this generalization, has suggested that the more important stories may be distinguished in terms of certain "formal variables." He cites as illustration of these distinguishing characteristics: the consistency of the story with other stories of this individual, its consistency with stories told by other persons to this picture, the faithfulness with which the subject followed instructions in telling this story, and, finally, the extent to which the individual "has perceived and apperceived the picture adequately in all its parts."

Indirect evidence bearing on this question is provided by the common observation that stories derived from particular cards frequently have little information pertinent to given variables. Thus the individual interested in studying aggression may find that stories told to certain TAT cards will only rarely provide pertinent evidence. Not only do the stories told by the subject vary in significance but *for given purposes* the stories characteristically evoked by different cards vary in significance. Eron (**19**, **20**) has shown that both incidence of themes and the emotional tone of stories differ significantly among the various TAT cards.

Although practical considerations and clinical experience indicate the necessity of an assumption of "crucial stories," there seems to be little empirical evidence that is directly pertinent.

Are Themes or Story Elements That Appear to Arise Directly out of the Stimulus Material Less Significant than Those That Are Not Directly Determined by the Stimulus Material?

The extent to which this assumption is embraced by projective testers is made clear in the general emphasis upon the importance of bizarre responses, sex reversals, etc. Rotter (**69**) gives it explicit stress in his interpretive discussion of the importance of "unusualness" of response in the Thematic Apperception Test. In the scoring of sentence completion responses, Rotter (**70**) has reported finding special diagnostic significance associated with "twist" or "reversal" endings to incomplete sentences. Weisskopf (**81**) has emphasized the assumption heavily in her suggestion that one method of estimating the diagnostic efficiency of projective stimulus material is through the use of a "transcendence index." This index is derived from the descriptive comments of the subject in responding to the stimulus material that go beyond "pure description."

The first difficulty we encounter in attempting to appraise this assumption is the question of how to go about measuring the degree to which a given response is determined by the stimulus material. This could be determined by estimating the structural similarities between the response and the stimulus. Or, one might develop empirical norms representing the common or usual response elements for each stimulus. Rosenzweig (**67**, **68**) has chosen the second alternative in his attempt to establish "apperceptive norms." The function of these norms, in large part, is to enable the interpreter to differentiate that which derives naturally from the stimulus card from that which is projective or personally determined. The norms reported by Eron (**19**) could be used in similar fashion. Presumably most clinicians implicitly employ both of these approaches. On the one hand they note responses which are an obvious denial of the observable elements in the stimulus material, and on the other they build certain expectations in regard to what a "normal" story is to each stimulus and they tend to place special interpretive emphasis upon departures from these norms.

Some projective testers probably object to this assumption on the grounds that "any" response, no matter how normal or directly derived from the stimulus material, may have significance in projective protocols. They might point to this attribute as one of the chief differentiae between mental testing (normative) and projective testing (idiosyncratic). It is certainly true that the extent to which a response deviates from the norm is much less a matter of concern to the interpreter of projective protocols than it is to the "mental tester." However, the fact remains that if a particular response is given by one hundred per cent of the persons taking the test, this response cannot possess diagnostic significance unless it is combined with some other response element that is less frequently encountered. In the latter case it is clear that the second or less frequent response is determining the interpretation, not the "normal" response. Nevertheless, it must be remembered that even a response given by a very high percentage of the respondents may have broad significance as in the case of a yes-no response to a specific question. In the limiting case, a response may be given by ninety-nine per cent of the respondents and still be an important diagnostic sign, especially in those few cases where it is not encountered.

Limited empirical justification for the view that significance and "stimulus-boundedness" are negatively related is provided in a study by McClelland, Burney, and Roby (49). They observed that *introduction* by the story-teller of an affiliated person into TAT stories where this person was not present in the picture was related to an experience the subjects had just undergone. A count of affiliated persons without consideration of whether or not they were in the picture did not reveal any relationship with the experimental treatment.

Supporting this assumption, we have the observation that most perception researchers accept the generalization that the more ambiguous the material, the easier it is to observe the operation of directive or motivational factors. Thus, they suggest that in those cases where the response bears relatively little relation to the stimulus material we are especially apt to observe dynamic or motivational factors. Closely related also is Bruner's (10) and Postman's (59) suggestion that the perceptual process may be represented as involving a relationship between the predetermining tendencies of the individual on the one hand, and information or stimulus constraints presented by the environment on the other. They imply that the stronger the determining tendency, the less in the way of environmental supports are needed to produce a related percept and the more contradictory information will be needed to produce a percept in opposition to the predetermining tendency. The above assumption is quite similar to Bruner's and Postman's view, since it proposes that the stronger the stimulus constraints or supports, the less we know about the predetermining tendencies in the individual that have produced a response consistent with these constraints. However, if the stimulus material is not consistent with the report of the subject, presumably the predetermining tendency was sufficiently strong to produce an appropriate response, even though the stimulus did not call for it.

Thus, the assumption that stimulus-bound responses are less diagnostic than responses that do not depend so heavily upon stimulus constraints appears to fit well with available empirical data.

Are Recurrent Themes in a Series of Stories Particularly Likely to Mirror the Impulses and Conflicts of the Story-Teller?

This is an assumption frequently stated but for which there is a minimum of pertinent empirical evidence. However, there are heavy rational considerations favoring such a view. The presence of the same theme even when the stimulus situation has been thoroughly altered implies strongly that there are impelling forces within the individual creating these themes rather than their being the inevitable outcome of stimulus constraints. The same theme following a change in stimulus material suggests that the response is not tied directly to, or produced by, a single specific stimulus. Thus, we may infer that the response tendency possesses generality and is more likely to be related to motivational factors than the response linked to a single stimulus. Even if we accept the view that in all instances the stimulus is evoking the response at question, we know that one measure of the strength of a drive is the extent to which stimulus generalization or displacement will occur. Thus, if the individual can equate a large number of stimuli in order to make this response, we may infer that the instigation lying behind the response is quite strong.

In addition, if interpretations are favored that incorporate a large amount of material, those based upon recurrent themes have a natural advantage in that the interpretation of one theme is automatically applicable to all of the other themes in the series. Finally, the presence of recurrent themes permits the investigator to sample or test the conditions under which these themes make their appearance.

In general, then, this assumption seems to have been accepted consistently by most projective testers although there is little or no direct evidence demonstrating its utility. However, there are a number of rational considerations that support such an assumption.

Do Stories Reflect not only Enduring Dispositions but also Momentary Impulses and Are These Both Reflected in the Same Manner?

Most investigators are primarily interested in the Thematic Apperception Test as a means of measuring enduring dispositions, although occasionally, especially in research, the measurement of situational or temporary instigations to behavior may be crucial.

There have been numerous studies showing the sensitivity of the instrument to temporary or situational determinants. Bellak (5) showed that the number of aggressive words in TAT stories increased when the story-teller was rebuked for the low quality of the stories he told. Sanford (71, 72) and Atkinson and McClelland (2) have shown that TAT protocols vary with food deprivation, and McClelland, Clark, and Roby (50) have shown story variation as a result of exposure to failure in a test situation. Rodnick and Klebanoff (66) have shown that TAT stories vary with relative success in a level of aspiration test. Lindzey (43) demonstrated that extrapunitive behavior on the part of the hero in TAT protocols increased significantly following failure in a social situation.

In similar fashion there are a number of studies that show the sensitivity of the stories to more enduring dispositions. It seems reasonable to accept individuals belonging to different psychiatric groups as differing in some enduring rather than situational attribute. Consequently, studies showing significant differences between groups separated on some diagnostic variable may be considered evidence of the test's sensitivity to enduring dispositions and conflicts. Balken and Masserman (3) observed significant differences in the TAT performance of patients categorized as conversion hysterics, anxiety hysterics, and obsessive compulsives. Renaud (64) was able with some difficulty to distinguish between psychoneurotics, traumatic brain disorder cases, and brain disease cases on the basis of TAT protocols. Cox and Sargent (16) found differentiating signs in the TAT performance of "stable" and "disturbed" school children. Working with mental hospital inmates Harrison (27) in approximately 77 per cent of the cases was able to identify accurately on the basis of TAT stories the psychiatric category in which the patient had been placed. He reports (28) similar results when "blind analysis" was employed, that is, when the tests were administered by a different person than the interpreter.

Further evidence of the sensitivity of the test to non-situationally determined motivation is supplied by studies such as Murray and Stein's (57) where a relatively high positive relationship was found between leadership ratings of ROTC candidates based on TAT performance and leadership ratings independently executed by officers of the men in question. Likewise, Henry's (30) study of the Navaho and Hopi suggests the ability of the test to discriminate these two groups, and further to supply psychological information concerning them that is consistent with information secured from extended observation or through the use of independent instruments. White (82) showed that there was a significant relationship between TAT response and the dispositions leading to hypnotizability. Harrison (27) in those cases where he was able to match descriptive statements based upon TAT responses with independent information derived from mental hospital case histories was correct in 82.5 per cent of the instances.

The question of whether temporary and enduring tendencies are reflected in stories in exactly the same fashion is important for two reasons. First, the clinician is primarily concerned with the more enduring tendencies and, therefore, it is desirable that he have some means of differentiating between these two classes of determinants. Second, some research, especially McClelland's (**2**, **50**) has implied that through studying the effect of situational factors upon TAT performance it is possible to arrive at means of interpreting TAT stories to reveal the operation of more enduring tendencies. If the reflection of these two kinds of motivational factors should prove to be very similar or the same process this would be unfortunate for the clinician, who would then be faced with the difficulty or impossibility of knowing whether a given tendency was temporarily instigated or whether this was a more permanent characteristic of the individual in question. On the other hand, such similarity in process would encourage experimental treatments as feasible means of approaching the task of specifying more exactly the means of inferring enduring tendencies.

McClelland (**47**) has supplied some evidence that the enduring and temporary processes are reflected in the same manner. He has demonstrated that the differences in subjects' TAT responses following experimental induction of a motivational state (threatening test situation) is related to other measures of the individuals' behavior, e.g., academic performance. This implies that the response of the subject to the immediate situation is related to his more permanent patterns of response.

McClelland and Liberman (**51**) derived a system for scoring need achievement in TAT responses that was based upon the differences produced in TAT protocols following experimentally induced failure in a situation related to achievement. In addition, they demonstrated that achievement as measured by this scoring technique was related to performance in an anagrams test and also related to the speed with which certain kinds of achievement related words could be recognized when exposed tachistoscopically. Thus, a measure of need achievement derived from a temporary instigation appeared to be related to other measures of achievement which were only distantly related to the initial instigating situation. While this evidence is not compelling, it does provide some support for the notion that temporary instigations affect TAT stories in a manner consistent with the way in which more enduring dispositions influence stories.

We appear to have excellent empirical evidence indicating that stories are responsive to both situational and enduring motivational factors. There is no conclusive evidence demonstrating the similarity or dissimilarity of the process whereby these two classes of determinants secure expression in the stories.

Do the Stories Reflect Events from the Past of the Subject That He Has Not Himself Actively Experienced? Are These Events Diagnostic of the Individual's Dispositions and Conflicts?

Inquiry following the customary administration of the TAT has verified the hypothesis that individuals do incorporate material taken directly from scenes that they have witnessed or from movies or books they have been exposed to. Murray (56) has reported this. Further, any clinician who has worked at all extensively with the TAT has inevitably many cases in his own experience of stories drawn directly from the world of the novel or drama, frequently with accompanying remarks indicating explicitly that this was the case. This is directly consistent with Freud's (24) early dictum that each dream incorporates something from the events of the preceding day.

Given the influence of these non-participated events, the question then becomes one of whether motivational factors affect or are revealed in the recall of such experiences. The selective function of memory and the importance of motivational determinants in the memory process have been recognized for some time, certainly since the appearance of Bartlett's (4) treatise on memory. A number of studies have investigated, under reasonably well-controlled conditions, memory as a function of such motivational variables as political ideology, Edwards (18), Levine and Murphy (41); value as measured by the *Study of Values*, McGinnies and Bowles (52); sex membership, Clark (14); mental set, Carmichael, Hogan and Walter (13); attitude, Postman and Murphy (60); punishment, McGranahan (53). Although there are many other factors, e.g., primacy, recency, vividness, known to influence recall, it seems reasonably well established that motivational factors do serve as one important class of determinants of memory.

If we accept these studies as evidence of the extent to which memory is influenced by motivational factors, it seems reasonable that the individual, in the process of recalling past events or experiences, will reveal or expose important aspects of his own motivational state. Consequently, the assumption that the particular events remembered by the subject are diagnostic of his dispositions and conflicts appears to be supported by available empirical data.

Do Stories Reflect Group-Membership, Cultural or Social Determinants as well as Personal or Individual Determinants?

This assumption simply implies consistent differences between the fantasy productions of individuals who belong to, or have been socialized in, different social groups. Thus, a certain amount of the variation in any TAT production can be accounted for by the fact that the individual has grown up in a given milieu or social role. The importance of the assumption derives from the fact that overlooking this kind of variation introduces a serious source of error in the interpretation of imaginative protocols from members of more than one social group.

Although little effort has been made to explore the variations in fantasy productions between many of the important groups of our own society, it is quite widely accepted or expected that these differences exist. Even such an important cleavage as that between male and female has been little explored so far as TAT behavior is concerned, while such variables as socio-economic status, occupational role, and ethnic group-membership have also been of slight interest to most investigators. Rosenzweig and Fleming (68) have reported differences between a roughly equated group of men and women on a number of specific aspects of TAT response. An investigation by Riess, Schwartz, and Cottingham (65) designed as a critical appraisal of the Thompson (78) modification of the TAT led to the observation of certain relatively slight variations in story length as a function of geographic residence and Negro-white group-membership.

Henry (30) in his investigation of Navaho and Hopi children found that inferences based upon his adaptation of the TAT related to independently secured information concerning the children and also that there were systematic differences between the Navaho and Hopi fantasy productions. In addition, he reports differences between Navaho subjects who were members of different subgroups in this society.

Although there is a paucity of empirical evidence demonstrating differences in fantasy production between various socio-cultural groups, what evidence is available appears to support this assumption.

Are Impulses and Conflicts Inferred from Stories Not Always Reflected Directly in Overt Behavior and Consciousness?

Projective testers vary considerably in the manner in which they emphasize this assumption. Some see a very intimate relation between imaginative behavior and overt behavior. Thus, Piotrowski introduces nine rules designed to permit the translation of fantasied into overt with the following statement:

> The rules proposed in this article are a new attempt to solve the problem of the relationship between the TAT and overt behavior. Since the TAT is mainly an exercise in creative imagination, it should reflect the patient's ideas and drives regardless of whether or not they find a direct expression in overt behavior. Thus, parts of the TAT always reflect the overt behavior of the subject while other parts reflect ideas which are not as directly manifested in overt actions. If this be so, we need a rule by means of which we could differentiate these two parts of the TAT. The rules presented below have been formulated largely for the purpose of meeting that need (58, p. 105).

Others are more cautious in stating their views of the relation between story-behavior and overt behavior. Murray, in his introduction to the TAT, suggests:

> It may be stated, as a rough generalization, that the content of a set of TAT stories represents second level, covert . . . personality, not first level, overt or public . . . personality. There are plenty of ways of discovering the most typical trends; the TAT is one of the few methods available today for the disclosure of covert tendencies. The best understanding of the total structure of personality is obtained when the psychologist considers the characteristics of manifest behavior in conjunction with the TAT findings . . . (**56**, p. 16).

In similar vein, Korner states:

> . . . instead of deploring the fact that fantasy and reality behavior do not necessarily correspond, as we currently seem to be doing, we can use projective techniques as a shortcut to a person's fantasy and ideational life, which then can be compared and examined in the light of his present and past actual behavior patterns (**36**, p. 627).

Although the relationship between covert and overt is assessed differently by various investigators, all seem to agree that the relationship is not perfect—fantasy behavior does not exactly mirror overt behavior. This omnipresent assumption can serve one of two functions depending upon the orientation of the investigator. It *can* serve simply as a convenient means of avoiding the necessity of ever being wrong. Thus, whenever inferences based on story protocols fail to relate to appropriate independent measures or observations, the clinician may simply point to the above assumption and add that only the naïve would expect always to observe linear relationships between imaginal and overt or conscious behavior. On the other hand, the investigator can use this assumption as a signpost pointing to one of the most important and difficult empirical problems facing the projective tester. This problem is the determination of the conditions under which inferences based on projective material directly relate to overt behavior and the conditions for the reverse.

It is possible to defend the position that projective techniques should not be expected to provide statements concerning overt behavior. Such a view implies that the techniques will always be used only as an "imaginal supplement" to an otherwise adequate description of the individual. Thus, given a person who "behaves" in a particular way, examination of his fantasy productions may permit us to make consistent or to account for behavior that hitherto was unaccountable. Certainly this represents an important function of these techniques. Equally certain is the fact that this is not the only circumstance under which these instruments are used. They *are* used as means of inferring overt behavior tendencies and presumably with more adequate rules of transformation they would be so used much more widely.

Investigations by Sanford *et al.* (**73**) and Symonds (**77**), which unfortunately from the point of view of sampling were both based on adolescent populations, demonstrate clearly that in some cases instead of the impulses inferred from TAT records being reflected in behavior, their converse or opposite appear in behavior. This observation leads to the question of whether impulses that secure release in overt behavior may not need to be expressed in fantasy productions. However, the many positive relationships between imaginal impulses and overt behavior in these studies and others make it clear that it is not an either-or proposition and that the statement of the actual conditions under which the impulse is revealed or concealed must be complex.

Sanford *et al.* (73) studied the relationship in a group of school children between fantasy ratings derived from the TAT and overt behavior ratings provided by teachers who had observed the children. They found an average correlation of +.11 between the two sets of ratings indicating clearly that the fantasy ratings alone were not good predictors of overt behavior. However, there were striking differences between the different variables used in the extent to which fantasy and behavior corresponded. For some needs there was a relatively high positive relationship, while for others there was a significant negative relationship between the overt and covert. In accounting for these findings, Sanford *et al.* suggested that those tendencies which were negatively sanctioned or prohibited would be high in fantasy and low in overt behavior, while those tendencies which were encouraged by society and for which the individual could secure complete overt expression would be high in behavior but low in fantasy. High ratings would be secured in both fantasy and overt behavior for those tendencies that society encouraged but did not permit complete freedom of expression in, e.g., achievement, dominance.

Murray (56) has suggested that tendencies not inhibited by cultural sanctions are apt to be highly correlated in their fantasy and overt expression. He reports a positive correlation of over .40 between fantasy and overt behavior for a group of college men on the following variables: abasement, creation, dominance, exposition, nurturance, passivity, rejection, and dejection. Negative correlations are reported between fantasy and overt for sex and no correlation between the two forms of expression for aggression and achievement. Korner (35) attempted to relate hostility as observed in a play situation with ratings of hostility in interpersonal relations with other children. She found no general relationship between the two sets of variables. Half of the children high on hostility in play situations were likewise high on hostility manifested in their dealings with other children. The remaining half who were high on hostility manifested in play situations were low on the second set of ratings. The investigator concluded that it was impossible to predict from the one situation to the other.

Symonds (77) related the fantasy themes of 40 adolescent boys and girls to adjustment ratings and teachers' ratings of behavioral characteristics. He concluded that the relationship between these two sets of variables was "insignificant and negligible."

Further evidence of the lack of a perfect relationship between fantasy and overt behavior is provided in those cases where an individual of known characteristics fails to reveal salient aspects of himself in his TAT constructions. Tomkins (80) reports the case of an individual who had a persistent spontaneous fantasy which included as an important theme a homosexual seduction. The TAT responses of this individual gave no sign of homosexual tendencies. In accounting for this and similar cases, Tomkins suggests that the important variables are the awareness of the subject of the impulse or tendency at question and the extent to which the tendency is condoned or accepted by society. If the impulse is known and unaccepted by society, the individual will prevent its appearance in the stories he tells. If he is unaware of the tendency, it will appear in his fantasy constructions even if it is negatively sanctioned by society. Bellak (6) reports several similar instances where TAT performance fails to reveal central aspects of the individual.

Relatively little has been done in the attempt to discover and formulate signs in the stories themselves that would provide evidence concerning the probability of overt expression. There is some evidence that, as Tomkins (79) proposes, the "psychological distance" maintained by the story-teller toward the impulse or disposition in question may be an important condition relating to the degree of overt translation.

Available empirical evidence clearly indicates that the assumed imperfect correlation between fantasied and overt behavior is warranted. However, at present, we are far from an adequate formulation of the signs or cues that might permit specification from fantasy protocols alone of the behavioral tendencies that will secure overt expression as opposed to those that will not.

RESEARCH IMPLICATIONS

Almost all of these assumptions point to further research that would be useful in clarifying their status. Perhaps more important than research aimed at further demonstrating the warranty of these same assumptions is research that attempts to provide a more exact statement of the conditions under which the assumptions are applicable and the way in which they can be related to empirical data. For example: what are the means by which the important story in a series can be determined? How can we determine whether or not a given fantasy impulse will receive overt expression? In what way do we determine whether or not a given response has been determined by the stimulus material? What are the circumstances under which symbolic transformations must be engaged in? How do we determine the empirical referent of a given symbol? Answers to these and a host of related questions are necessary before we can hope to provide the TAT user with an explicit, repeatable set of operations for inferring motivational states.

In addition to problems connected with the interpretive assumptions and the more specific questions implied in the above paragraph, there is also the matter of formulating explicitly a method of scoring TAT protocols that is practical, intersubjective, and able to embrace a reasonable number of the behavioral variables in common use. To a large extent standardization and specific interpretive rules must wait until some agreement has been reached by most TAT users as to the major aspects of TAT response that will be focused upon in analysis.

SUMMARY

This paper has stressed the continuity between projective testing and research and theory in other areas of psychology. In addition, it has presented ten assumptions commonly made in interpreting Thematic Apperception Test stories and has sampled briefly the empirical evidence that can be used to confirm or reject each of them.

BIBLIOGRAPHY

1. Arnold, Magda B. A demonstration analysis of the TAT in a clinical setting. *J. abnorm. soc. Psychol.*, 1949, **44**, 97–111.
2. Atkinson, J. W., & McClelland, D. C. The projective expression of needs: II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.
3. Balken, Eva R., & Masserman, J. H. The language of fantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state, and obsessive-compulsive neuroses. *J. Psychol.*, 1940, **10**, 75–86.
4. Bartlett, F. C. *Remembering.* London: Cambridge Univ. Press, 1932.
5. Bellak, L. The concept of projection: an experimental investigation and study of the concept. *Psychiatry*, 1944, **7**, 353–370.
6. Bellak, L. Thematic apperception: failures and the defenses. *Trans. N. Y. Acad. Sci.*, 1950, **12**, 122–126.
7. Blake, R. R., & Wilson, G. P. Perceptual selectivity in Rorschach determinants as a function of depressive tendencies. *J. abnorm. soc. Psychol.*, 1950, **45**, 459–472.
8. Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
9. Bruner, J. S. Perceptual theory and the Rorschach Test. *J. Personality*, 1948, **17**, 157–168.
10. Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. Ramsey (Eds.), *Perception an approach to personality.* New York: Ronald Press, 1951. Pp. 121–147.
11. Bruner, J. S., & Postman, L. Tension and tension release as organizing factors in perception. *J. Personality*, 1947, **15**, 300–308.
12. Bruner, J. S., & Postman, L. An ap-

12·8·70
LIBRARY

proach to social perception. In W. Dennis (Ed.), *Current trends in social psychology*. Pittsburgh, Pa.: Univ. Pittsburgh Press, 1948. Pp. 71–118.

13. Carmichael, L., Hogan, H. P., & Walter, A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.
14. Clark, K. B. Some factors influencing the remembering of prose materials. *Arch. Psychol., N. Y.*, 1940, No. 253.
15. Cofer, C. N., & Foley, J. P. Jr. Mediated generalizations and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
16. Cox, Beverly, & Sargent, Helen. TAT responses of emotionally disturbed and emotionally stable children: clinical judgment versus normative data. *Rorschach Res. Exch.*, 1950, **14**, 61-74.
17. Diven, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.
18. Douglas Anna G. A tachistoscopic study of the order of emergence in the process of perception. *Psychol. Monogr.*, 1947, **61** (6), Whole No. 287.
19. Edwards, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36**, 34–50.
20. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, (9), Whole No. 315.
21. Eron, L. D., Terry, Dorothy, & Callahan, R. The use of rating scales for emotional tone of TAT stories, *J. consult. Psychol.*, 1950, **14**, 473–478.
22. Farber, L. H., & Fisher, C. An experimental approach to dream psychology through the use of hypnosis. *Psychoanal. Quart.*, 1943, **12**, 202–216.
23. Franck, Kate. Preferences for sex symbols and their personality correlates. *Genet. Psychol. Monogr.*, 1946, **33**, 73–123.
24. Franck, Kate, & Rosen, E. A projective test of masculinity-femininity. *J. consult. Psychol.*, 1949, **13**, 247–256.
25. Freud, S. The interpretation of dreams. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 181–552.
26. Grandine, Lois, & Harlow, H. F. Generalizations of the characteristics of a single learned stimulus by monkeys. *J. comp. physiol. Psychol.*, 1948, **41**, 327–338.
27. Haggard, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *J. exp. Psychol.*, 1943, **33**, 257–284.
28. Harrison, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients: II. A quantitative validity study. *Character & Pers.*, 1940, **9**, 122–133.
29. Harrison, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients: III. Validation by the method of "blind analysis." *Character & Pers.*, 1940, **9**, 134–138.
30. Harrison, R., & Rotter, J. B. A note on the reliability of the Thematic Apperception Test. *J. abnorm. soc. Psychol.*, 1945, **40**, 97–99.
31. Henry, W. E. The Thematic Apperception Technique in the study of culture-personality relations. *Genet. Psychol. Monogr.*, 1947, **35**, 3–315.
32. Hull, C. L. *Principles of behavior: an*

introduction to behavior theory. New York: Appleton-Century, 1943.

33. Hull, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.
34. Katz, I. Emotional expression in failure: A new hypothesis. *J. abnorm. soc. Psychol.*, 1950, **45**, 329–349.
35. Klein, D. B. The experimental production of dreams during hypnosis. *Univ. Texas Bull.*, 1930, No. 3009.
36. Korner, Anneliese F. *Some aspects of hostility in young children*. New York: Grune & Stratton, 1949.
37. Korner, Anneliese F. Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, **45**, 619–627.
38. Krout, Johanna. Symbol elaboration test: The reliability and validity of a new projective technique. *Psychol. Monogr.*, 1950, **64** (4), Whole No. 310.
39. Lashley, K. S., & Wade, Marjorie. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
40. Levine, R., Chein, I., & Murphy, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283–293.
41. Levine, J. M., & Murphy, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38**, 507–517.
42. Levy, D. M. Projective techniques in clinical practice. *Amer. J. Orthopsychiat.*, 1949, **19**, 140–144.
43. Lindzey, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296–309.
44. MacCorquodale, K., & Meehl, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
45. Mayman, M., & Kutner, B. Reliability in analyzing Thematic Apperception Test stories. *J. abnorm. soc. Psychol.*, 1947, **42**, 365–368.
46. McCleary, R. A., & Lazarus, R. S. Autonomic discrimination without awareness: an interim report. *J. Personality*, 1949, **18**, 171–179.
47. McClelland, D. C. Measuring motivation in phantasy: the achievement motive. In H. Guetzkow (Ed.), *Groups, leadership and men: Research in human relations*. Pittsburgh: Carnegie Press, 1951.
48. McClelland, D. C., & Atkinson, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205–222.
49. McClelland, D. C., Burney, R. C., & Roby, T. B. The effect of anxiety on imagination. Paper read at Eastern Psychological Assn., 1950.
50. McClelland, D. C., Clark, R. A., Roby, T. B., & Atkinson, J. W. The projective expression of needs: IV. The effect of need for achievement on thematic apperception. *J. exp. Psychol.*, 1949, **39**, 242–255.
51. McClelland, D. C., & Liberman, A. M. The effect of need for achievement on recognition of need-related words. *J. Personality*, 1949, **18**, 236–251.
52. McGinnies, E., & Bowles, W. Personal values as determinants of perceptual fixation. *J. Personality*, 1949, **18**, 224–235.
53. McGranahan, D. V. A critical and experimental study of repression. *J. abnorm. soc. Psychol.*, 1940, **35**, 212–225.
54. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

55. Murray, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310–329.
56. Murray, H. A. *Thematic Apperception Test manual.* Cambridge: Harvard Univ. Press, 1943.
57. Murray, H. A., & Stein, M. I. Note on the selection of combat officers. *Psychosom. Med.*, 1943, **5**, 386–391.
58. Piotrowski, Z. A. A new evaluation of the Thematic Apperception Test. *Psychoanalyt. Rev.*, 1950, **37**, 101–127.
59. Postman, L. Toward a general theory of cognition. In J. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harpers, 1951.
60. Postman, L., & Murphy, G. The factor of attitude in associative memory. *J. exp. Psychol.*, 1943, **33**, 228–238.
61. Rapaport, D. The clinical application of the Thematic Apperception Test. *Bull. Menninger Clin.*, 1943, **7**, 106–113.
62. Razran, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.
63. Razran, G. Attitudinal determinants of conditioning and of generalization of conditioning. *J. exp. Psychol.*, 1949, **39**, 820–829.
64. Renaud, H. Group differences in fantasies: head injuries, psychoneurotics, and brain diseases. *J. Psychol.*, 1946, **21**, 327–346.
65. Riess, B. F., Schwartz, E. K., & Cottingham, Alice. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700–709.
66. Rodnick, E. H., & Klebanoff, S. G. Projective reactions to induced frustrations as a measure of social adjustment. *Psychol. Bull.*, 1942, **39**, 489. (Abstract)
67. Rosenzweig, S. Apperceptive norms for the Thematic Apperception Test: I. The problem of norms in projective methods. *J. Personality* 1949, **17**, 475–482.
68. Rosenzweig, S., & Fleming, Edith. Apperceptive norms for the Thematic Apperception Test: II. An empirical investigation. *J. Personality*, 1949, **17**, 483–503.
69. Rotter, J. B. Thematic apperception tests: suggestions for administration and interpretation. *J. Personality*, 1946, **15**, 70–92.
70. Rotter, J. B., Rafferty, Janet E., & Schachtitz, Eva. Validation of the Rotter Incomplete Sentences Blank for college screening. *J. consult. Psychol.*, 1949, **13**, 348–355.
71. Sanford, R. N. The effects of abstinence from food upon imaginal processes: a preliminary experiment. *J. Psychol.*, 1936, **2**, 129–136.
72. Sanford, R. N. The effects of abstinence from food upon imaginal processes: a further experiment. *J. Psychol.*, 1937, **3**, 145–159.
73. Sanford, R. N., Adkins, Margaret M., Miller, R. B., *et al.* Physique, personality and scholarship: a cooperative study of school children. *Monogr. Soc. Res. Child Developm.*, 1943, **8**, No. 1.
74. Sears, R. R. Effects of frustration and anxiety on fantasy aggression. *Am. J. Orthopsychiat.* (In press.)
75. Siipola, Elsa M. The influence of color on reactions to ink blots. *J. Personality*, 1950, **18**, 358–382.
76. Stein, M. I. Personality factors involved in the temporal development of Rorschach responses. *Rorschach Res. Exch.*, 1949. **13**, 355–414.
77. Symonds, P. M. *Adolescent fantasy: An investigation of the picture-story method of personality study.* New York: Columbia Univ. Press, 1949.
78. Thompson, C. E. The Thompson modification of the Thematic Ap-

perception Test. *Rorschach Res. Exch.*, 1949, **13**, 469–478.

79. TOMKINS, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.
80. TOMKINS, S. S. The present status of the Thematic Apperception Test, *Am. J. Orthopsychiat.*, 1949, **19**, 358–362.
81. WEISSKOPF, EDITH A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379–390.
82. WHITE, R. W. Prediction of hypnotic susceptibility from a knowledge of subjects' attitudes. *J. Psychol.*, 1937, **3**, 265–277.
83. WICKENS, D. D. Stimulus identity as related to response specificity and response generalization. *J. exp. Psychol.*, 1948, **38**, 389–394.
84. WYATT, F. Formal aspects of the Thematic Apperception Test. *Psychol. Bull.*, 1946, **39**, 491. (Abstract)

Received May 1, 1951.

WHEN NOT TO FACTOR ANALYZE

J. P. GUILFORD
University of Southern California

The apparent increase in the utilization of factor-analytical methods is undoubtedly gratifying to all those who have championed those methods. Their application to research in clinical psychology as well as to research in what has traditionally been known as experimental psychology speaks well for their versatility.

It seems desirable, however, to interject some words of caution, in view of the number of misuses of factor analysis that are appearing from time to time in published articles. Many a report that includes an account of a factor analysis is faulty because of failure to take certain precautions to assure an adequate solution by that approach.

In some instances there have been poor choices of experimental variables or of populations, or of both. Sometimes it seems as if the investigator, feeling the urge to factor analyze, applies the method to his next investigation in which intercorrelations are conveniently available. Too many experimental variables have been analyzed just because they are conveniently at hand. More specifically, scores from such sources as the *Strong Vocational Interest Blank*, the *Kuder Preference Record*, Bernreuter's *Personality Inventory*, the *Minnesota Multiphasic Personality Inventory*, and the Guilford-Martin personality inventories are inappropriate experimental variables to use under most conditions of analysis. Why these variables are inappropriate will be explained in what follows.

If any research tool is to give meaningful results, it must be applied in the right place and in the proper manner. Writers on factorial methods have expressed warnings from time to time concerning requirements and limitations for the use of those methods. The most common misuses of those methods, however, cannot be entirely excused on the grounds of unfamiliarity with the literature on factor-analysis techniques. Many of the misuses and abuses could have been avoided if the investigators had observed the ordinary good rules of experimental controls and of population sampling.

The use of a complicated statistical procedure like factor analysis does not permit one to forget about the usual safeguards that should surround scientific observations. Statistical operations do not compensate for carelessness in making observations. Rather, they presuppose careful observations. They then serve as an important aid in seeing order in the observations and in making sense of that order. Under inappropriate conditions of observation, data may appear to have an order that is misleading if not fictitious. There is no statistical magic that will give a good ordered view of nature when the data do not permit.

The discussion here will not attempt to include all the errors and pitfalls that have occurred in connection with factor analyses. In the first place, the context is limited to those factorial methods in which rotation of reference axes is an important feature. It is assumed that some rotational procedure is usually necessary to achieve an order in data that has a parallel in psychological concepts. In other words, the bias is in favor of the principle that rotations of reference axes are generally needed to yield a meaningful reference frame. It is also assumed that the order underlying a number of experimental variables is substantially simpler than the data from which it was extracted. In other words, the number of common factors is definitely less than the number of experimental variables. When the term "factor" is used hereafter, the idea of "common factor" is implied.

Some Common Faults in Factor Analysis

1. *Too many factors are often extracted for the number of experimental variables.* Too many factors for the number of experimental variables is likely to preclude a good rotational solution. Thurstone has repeatedly emphasized that in a rotational solution the position of each reference axis must be overdetermined (8). A good rule is to have at least three variables for every factor expected. If there has been good preparatory planning of the study, one can predict the probable number of factors and whether the variables are well distributed among the factors. A paucity of variables is not only a handicap in making rotations but also in the interpretation of factors. It requires several tests all substantially loaded with a factor to indicate the generality of the factor and to define it satisfactorily.

2. *Too many experimental variables are factorially complex.* Rotations and interpretations would be much simplified if each variable were of complexity one; that is, if it measured only one common factor to any appreciable extent. This is an ideal that we achieve in test construction only once in many attempts. There is little excuse for taking almost any variable that is handy. Such variables, where there has been no effort to restrict them, are very likely to measure two or more common factors. With appropriate planning and careful test construction we can do much to reduce the complexity of tests in preparation for an analysis. The effort will pay off in terms of facilitated rotations and interpretations. It may make the difference between success or failure in achieving a solution that is acceptable.

3. *Sometimes a common factor fails to come out because it is substantially represented in only one experimental variable.* For a common factor to emerge in a particular set of experimental variables, more than one variable must have substantial variance in that factor.[1] It is otherwise left as specific variance so far as that particular analysis is concerned. This is ordinarily not a serious matter. It may merely mean that an opportunity has been missed to find a certain factor that is actually present.

But this is sometimes not the whole story. If this factor should be represented in minor amounts in some other variables, this common-factor variance may be extracted. But there is likely to be trouble in making rotations. If the rotating is done entirely blindly (without regard to known landmarks or psychological meaning) such a factor is frequently lost. Its variance may be divided several ways.

A recent example of this occurred in analyzing a battery of reasoning tests. A test of numerical operations had been put in the battery to assure the segregation of the number-factor variance expected in some of the reasoning tests. The number factor is so well established and a numerical-operations test is such a unique measure of it, that any solution not preserving this factor would be very suspect. Only the insistence of an axis through the numerical-operations test preserved the factor. This undoubtedly led to a clearer picture with respect to the reasoning factors.

This same fault comes up in a similar manner in connection with the analysis of scores from some personality inventories. Where some of the scores are fairly unique for single factors, those factors either do not come out in the analysis, or if weakly represented in other scores the rotational solution tends to break up their variances. This is particularly true when fewer common factors are extracted than actually are present in the scores.

4. *Not enough factors are extracted.* Experience has shown that it pays to extract a liberal number of factors. Most criteria for deciding how many factors to extract are not sufficiently liberal. No harm appears to be done in extracting too many factors. If more than the proper number of factors are extracted, the rotations will lead to the rejection of one or more of them as "residual" factors. The presence of one or two extra factors often helps to clear up the general picture of the structure for the factors that are of genuine significance.

Some investigators seem to fear that the last factors extracted are entirely a matter of error variance. It is as if they believe that up to a certain point in the extractions all variance extracted is common-factor variance and after that all variance extracted is error variance. The fact that the late-extracted variance helps to clear up the structure in rotations and to improve psychological meaningfulness of the factors is some evidence that true variance may be extracted late as well as early. There seems to be no good reason for rejecting the idea that error variance is also extracted early as well as late.

---

[1] This assumes the usual practice of using communalities in the diagonal cells of the correlation matrix.

From this discussion it would seem to follow that the best criterion of when to cease extracting factors is the size of the larger factor loadings. As long as they are large enough to contribute something to other factors in rotation or to be built up into something psychologically meaningful they probably should have been extracted.

5. *Correlation coefficients used in analysis are often spurious.* Thurstone has often emphasized the point that the correlations used in factor analysis should be between variables that are linearly independent (8). This means that there should be no reason for covariation except that due to common factors.

There are a number of situations in psychological investigations in which specific and error variances actually contribute to intercorrelations where they should not be permitted to do so. One common situation is in connection with personality inventories in which the same items are scored with weights for more than one trait variable. This is especially true of the Bernreuter *Personality Inventory*, the *Strong Vocational Interest Blank*, and the Guilford *Inventory of Factors STDCR*. It is true to some extent of the *Guilford-Martin Inventory of Factors GAMIN*. For every item that is weighted in two scores there is a contribution to the obtained correlation between those scores. In so far as the item measures factors in common to the two scores this is legitimate. But the item's contribution to any total score also includes some specific and error variance. The specific and error variances thus contribute to the intercorrelations of items.

Let us assume that two scores were designed to measure two factors that are actually orthogonal or independent. But in the two scoring keys there are a number of items weighted similarly. There would consequently be a positive correlation, perhaps a substantial one, between scores in these two variables. A negative correlation could also be brought about between scores for two independent factors if the weights were in opposite direction when items were scored for both factors. The positive and negative intercorrelations among scores on the *Strong Vocational Interest Blank* are largely influenced by these positive and negative correlations of weights. To what extent the intercorrelations of multi-scored inventories represent actual degrees of relationship of the variables the test author intended to measure and to what extent they represent these incidental communities of specific and error variances is unknown.

The writer is often asked why some of the scores on his inventories intercorrelate so strongly when the scores were designed to measure factors. This question confuses two different things; on the one hand the factors, and on the other hand the scores which were designed to measure those factors. Factors and their corresponding scores are logically and operationally distinct variables.

There are several reasons for the correlation of factor scores. One is the fact of intercorrelations of the factors themselves. It is very probable that some of the temperament and interest factors are actually interrelated. It is the writer's belief that at the present time we are not in a very good position to determine the extent of those intercorrelations satisfactorily. Just as there are adventitious conditions that distort the intercorrelations of experimental variables, there are also incidental disturbing conditions that give the appearance of intercorrelations of factors in an oblique solution in a factor analysis.

In addition to the genuine intercorrelations of factors themselves, there is the overlap of specific and error variances mentioned above contributing to the correlations among the scores of the factors. There is also the fact that items themselves are factorially complex. The items designed to measure factor *A*, on the whole, may be frequently involved with factor *B* and items designed to measure factor *B* may be involved with factor *A*. The factors *A* and *B* might be orthogonal but under the present level of skills in item writing it may be difficult to effect a separation in the items written to measure either factor. Even though no itém were weighted for both factors in scoring, both scores would carry variance in both factors. The solution may be in the use of suppression variables (6). We should have to make some assumption about the actual factor intercorrelations, however, to assure by this method a degree of interrelation of scores consistent with interrelation of factors.

If a set of scores for factors were properly slanted so that each score measures one factor and one only, the intercorrelations of those scores would represent the intercorrelations of the factors. The factor analysis of these intercorrelations would give the second-order factors. When the intercorrelations of factor scores are distorted by all the sources of error that have just been described, however, it can be seen that an analysis will yield results that are difficult to interpret, at best.

6. *Correlations of ipsative scores are sometimes used in an analysis.* The term "ipsative measurement" was emphasized by Cattell (2). Ipsative measurements can best be defined by contrast to the more common "normative measurements." In normative measurements there is a scale for every *trait* and a population of individuals is distributed about the mean of that population. In ipsative measurements, there is a scale for every *individual* and a population of an individual's trait scores is distributed about that individual's mean. Normative scores are used to indicate inter-individual differences in a trait. Ipsative scores are used to indicate intra-individual differences in a number of traits. We compare individuals with respect to a trait by using normative scores. We compare trait quantities within an individual by using ipsative scores.

For the usual factor analysis (*R*-technique) in which the experimental variables are individual differences, normative scores are properly used. We correlate scores over a population of individuals. When the factor analysis is by the *Q*-technique, in which individuals are intercorrelated, we should use ipsative scores. The scores are then correlated over a population of traits or qualities. It is improper to use normative scores in a *Q*-technique analysis and to use ipsative scores in an *R*-technique analysis. Unfortunately, both of these mistakes are sometimes made.

One of the best examples of ipsative measurements in common practice is found in the scores from the *Kuder Preference Record*. Though these scores are not entirely ipsative, they partake strongly of the properties of ipsative measurements. The reason for this lies in the forced-choice type of item used. Where interest variables are pitted against one another rather systematically for preferential judgments, the result is that a high preference for any one interest to that extent means low preferences for other interests. It is consequently impossible for an individual to score very high in all, or nearly all, of the Kuder interest variables, as should be possible in normative scoring.

The typical profile of an individual on the Kuder inventory is distinguished by the unusual number of extreme scaled scores, high and low. This is approved by the counselor, who wants to high-light differences among the interest traits. But it represents a kind of measurement that is not suited to intercorrelations of the usual kinds where normative measurements are needed. The correlations of the Kuder scores with each other and with outside variables are of questionable meaning to the extent that they are ipsative in nature. The effect on the intercorrelations of the Kuder scores is that about two-thirds of them are negative (**5**, p. 615).

The scores on the Strong interest inventory have some of the ipsative quality about them, in view of the preferential items in some parts of that inventory. The effects of these parts are superimposed upon the effects of correlations of weights already mentioned.

Scores based upon the complete paired-comparison presentation of items are even more completely ipsative. The means of all individuals in the trait scores involved are equated, which would not be true in normative measurements. Such scores would be appropriate for analysis by the *Q*-technique, if the score sample were sufficiently large, but not by the ordinary *R*-technique.

7. *A pair of factors is very much confined to the same experimental variables.* This is not a very common circumstance, but one for which one should be on the alert. If factors *A* and *B* are commonly measurable by the same tests, so that in the battery being analyzed no test having variance in the one factor is free of variance in the other, factors *A* and *B* may be difficult to separate. This is especially true where an insufficient number of factors has been extracted. Sometimes one can

detect the fact that two well-known factors have thus "telescoped" into one. But how often this happens with factors that are new or unknown is hard to say. The difficulty can be forestalled to some extent by care in the selection and the construction of tests to be analyzed.

8. *The population on which the analysis is based is heterogeneous.* Although no tests of statistical significance are usually made in connection with a factor analysis, the common sampling problems have a bearing upon the success of the analysis. Too often, samples are taken from populations that happen to be most available. What is worse, different populations are thrown together without questioning what effect this may have upon the intercorrelations.

Although a number of studies have shown that the same factors may be found in the same tests when analyzed in somewhat different populations (e.g., white vs. Negro, male vs. female) with somewhat similar factor loadings (**3**, **6**), the combining of populations is a different matter. If there are notable differences between means in both of two variables, there is a correlation between the means of those two variables. Correlations between means influence correlations of scores when such populations are thrown together.

One should not pool the data derived from different populations for the purpose of computing intercorrelations, then, unless he can show that differences in population means on the experimental variables are insignificant or unless he makes allowances for any "between-means" correlations. One procedure would be to compute the correlation matrix for each population separately. If the corresponding coefficients are similar they may be averaged to obtain a single matrix for analysis. The chief external conditions on which homogeneity should ordinarily be achieved are the familiar ones of age, sex, and education, wherever it is suspected that any of the experimental variables are related to those conditions.

9. *Not enough attention is given to requirements for correlation coefficients.* A factor analysis ordinarily begins with correlation coefficients. The results of the analysis can be no better than the data with which the analysis begins. Even elementary statistical textbooks mention the conditions needed in order that the correlation in a sample shall be a good estimate of the correlation in the population. Yet, it is apparent that many who analyze pay little attention to the question of whether those conditions are satisfied. At least, in their reports very little is said about the satisfaction of those conditions.

The least that the investigator can do is to examine the form of the frequency distribution of each variable that is correlated. The distributions should not be markedly skewed, truncated, or multi-modal. Any of these departures from unimodal, complete, and symmetrical distributions can endanger proper estimates of population correlations. The departures would probably have to be severe enough as to be obvious by inspection in order to justify concern. In other words, it would not require statistical tests of such departures to tell whether one should do something about them. If all distributions are fairly unimodal and symmetrical, the chances are that regressions are rectilinear and that homoscedasticity would prevail in the bivariate plots.

One cannot depend upon standard psychological tests to yield the regular forms of distributions just specified for computing Pearson product-moment *r*'s. Distributions vary as a test is administered to different populations. The form of distribution for new and untried tests, particularly, is likely to be irregular unless pains have been taken by pre-testing to assure regular distributions.

The remedies for irregular distributions, should they occur, are common knowledge. Skewed distributions can be normalized to some standard scale, provided the raw-score scale has considerable range. They cannot be satisfactorily normalized if there are too few raw-score categories or if there is truncation. For a distribution with undue numbers of cases in end categories there is only one very convenient remedy—to dichotomize the distribution.

Dichotomizing a distribution means the computing of biserial *r*'s with other variables that are not dichotomized. Since biserial *r*'s are fairly good estimates of Pearson *r*'s, they may be mixed with the Pearson *r*'s for the analysis. Some investigators hesitate to dichotomize a distribution for computing a biserial *r* when the sample distribution is skewed or truncated, in view of the assumption of normality that is made about the dichotomized variable. The assumption of normality underlying the computation of a biserial *r* is that the *population* distribution is normally distributed. The population distribution can be normal even though the sample distribution is not, due to a faulty measuring scale. If there is no decisive information to the contrary, the population distribution on a psychological variable may often be assumed to be normal.

If more than one distribution is dichotomized, the coefficient to use is the tetrachoric *r* for a pair of such distributions. If the sample is fairly large (at least approximately 400) it would be convenient and defensible to dichotomize all the distributions. Whatever dichotomizing is done should divide distributions as near the medians as possible, in order to minimize the standard errors of the coefficients. Dichotomizing distributions near the medians has another virtue in connection with factor analysis. It tends to equate difficulty levels of the tests. It has been suggested that comparable difficulty levels should be achieved in order to avoid distortions in correlation coefficients (**1**).

10. *Difficulty levels of tests often vary substantially.* It is well known that variations in difficulty level of tests relative to the ability level of the population affect the form of distributions of scores. Inappropriate difficulty level is followed by skewing, and skewing, as was pointed out above, is followed by distortions in correlation coefficients. Certain remedies for skewing were pointed out above, but they are only correc-

tions after the fact. Scaling and dichotomizing in order to eliminate the effects of skewing appear to take care of the biases due to inappropriate difficulty levels, but they may not do so completely.

If we dichotomize a seriously skewed distribution for a very difficult test, have we effected the same kind of division of the sample as we would have done with an easier test? We do not know. We do know that a symmetrical distribution makes better use of the available range of scores. If we have a relatively short test of 20 to 25 items, for example, and if the distribution is skewed, the range of utilized score categories becomes very much restricted. With a symmetrical distribution there is opportunity for a larger range and a finer scaling.

It follows that we should attempt to avoid skewed distributions from the start. This can be achieved by good test construction and by selecting tests with optimal adjustment of difficulty level to the status of the population. If we assure regular distributions, we may then compute the more reliable Pearson r's and avoid the recourse to scaling and dichotomizing or other remedial measures.

HOW THE FAULTS APPLY TO STANDARD TESTS

It was stated earlier that the scores from certain personality inventories are not suitable variables for use in factor-analysis investigations. The reasons are now much more easily presented in terms of the principles just discussed.

The scores on the *Strong Vocational Interest Blank* are not good variables for analysis because they entail the faults numbered 1, 2, 5, and 6. Covering the scope of vocational interests (also some temperament qualities that are latent in the scores) as they do, the scores may be no more numerous than the factors involved. Most of the scores are probably factorially complex, since the distinguishing qualities of a vocational group are numerous. No attempt was made to seek score categories that are functionally unique. The intercorrelations of the scores are quite generally influenced by the multi-scoring technique which contributes identical specific and error variances to two or more scores. The use of some forced-choice items introduces some ipsative properties into the scores, though this feature is probably of relatively small importance in the Strong inventory.

The *Kuder Preference Record* scores are not suitable variables for factor analysis by the usual R-technique because of faults 1, 3, and 6. Among the limited number of Kuder scores there may be more factors than scores. It is possible that one or more of the Kuder scores actually approach univocality for primary interests. If so, within the context of the Kuder score variables, these scores would yield no common factors in an analysis. If other variables were brought into the picture those factors might emerge. The strong ipsative property of the Kuder scores, however, renders their use for intercorrelations among themselves and with other variables so questionable as to preclude attempts at analysis by the *R*-technique.

The analysis of the Guilford-Martin inventory scores by themselves in the search for first-order factors is of little use because of faults 1, 3, and 5. From previous factorial analyses of single items the indications are that there are 13 distinct factors involved. For the most part, each factor is strongly represented in only one score. For the *Inventory of Factors STDCR*, there are seriously biased scores due to multiple scoring of items. This is a minor feature of the *Inventory for Factors GAMIN*. In studies which include other variables in which any of these 13 factors are expected, it would be well to use the appropriate scores from these inventories as reference variables. Even then, it has been the policy of the writer not to include some combinations of these scores in the same analysis but to determine the correlation of a new factor with these scores after the analysis has been completed. Another procedure has been to break up the items scored for any one factor into two or three groups, each more homogeneous in appearance than the total list, and to include the part scores in the analysis. The multiple scoring of items is avoided in this procedure.

Analysis of Bernreuter's *Personality Inventory* and of the *Minnesota Multiphasic Personality Inventory*, with total scores as variables, is precluded by reason of faults 1, 2, and 5. Neither are these same scores helpful in connection with analyses of other scores because of their probable complexity and their unknown factorial compositions.

Applications of factor analysis to scores from the Rorschach instrument also encounter to a marked degree some of the difficulties mentioned here, as well as others that have not been mentioned. Many of the customary Rorschach scores are linearly dependent and distributions are usually irregular.

AVOIDING FAULTY STEPS IN FACTOR ANALYSIS

Ways of avoiding the faults discussed above are fairly obvious. Some of them have been mentioned in discussing the faults. It is perhaps of more value to stress good principles of "hygiene" for factor analysis than to dwell upon errors.

Requirements for effective factorial investigations have been mentioned more than once in the literature (4). Of all the general policies that might be mentioned, the one that seems to be in need of emphasizing most is that considerable careful planning should precede an analysis. It is in the designing of the particular study and later in the interpreting of factors that the psychologist can really show his skill as a psychological investigator. He should enjoy those aspects of his investigation most. What happens between the planning and interpretation stages can be carried out by laboratory assistants and clerks. If the genuinely psychological aspects of the investigation are not conducted with thoughtfulness and care, no amount of computational activity will make up for those deficiencies.

Briefly, a factorial study should begin with a decision as to the scope of the domain to be investigated. If one does not wish to run the risk of being involved with an unwieldy list of experimental variables, he will find it wise to select a rather limited domain for a single analysis. A domain that is not likely to involve more than 15 factors or 50 experimental variables is good policy. It is not necessary to analyze for all common factors simultaneously. The structures of the factors from limited studies can be made to fit together to complete the larger picture. There is also much to be gained from repeating analyses with variations and with combinations of domains.

The initial planning should emphasize the formation of hypotheses as to what factors are likely to be found in the selected domain and as to the probable properties of such factors. There should be no hesitation to indulge in this type of activity. Perhaps some of the more rigorously inclined investigators have avoided this approach because they have seen some of their less rigorous colleagues "dream up" hypothetical traits to accept them as established fact. Explicit hypothesis formation is something of which psychologists have done too little in their planning of research. It should make for more substantial progress for the factor analyst to begin an investigation by asking "May I assume the existence of factor *X*?" or "Does factor *Y* have these properties, or these, or these?" With properly designed experimental variables, the answer from the analysis should be of the form, "Yes, you may," or "No, you may not assume the existence of factor *X*"; "Factor *Y* is more like this than it is like that." Some investigators may be afraid of the cliche "You get out of a factor analysis exactly what you put into it." The experienced analyst only wishes that he could come closer to achieving this status of omniscience and control!

Tests especially constructed for a particular study are likely to give much clearer answers to the hypotheses than are tests already available, except where there is previous knowledge of factors in those tests. It is important for each test to be as homogeneous functionally as we can make it. It is desirable to select items for internal consistency. The usual item-analysis procedures will not assure the construction of a one-factor test. The scale-analysis methods of Guttman might be helpful in achieving one-factor tests. The cumbersomeness of the procedure makes the cost rather prohibitive, however, and the end result may be a score that measures something entirely too specific. The best practical recourse is to the psychologist's skill in defining a hypothesized factor, in writing items, and in editing them. Ordinary item-analysis procedures will be of material help in assuring a large proportion of true variance in the scores.

In the selection of experimental variables, it is important to take into account all variances in already known factors. If one cannot exclude all common variances in known factors that he wants to leave out of the picture, it becomes necessary to put in the battery a good measure or two for each such factor. This not only identifies the factor but also segregates its variance so that it does not muddy the waters with respect to the rest of the factor structure.

With these general suggestions, most of which are not new, in mind, and with observance of the suggestions expressed or implied in connection with the various faults discussed in this paper, the one who would factor analyze will be on the road to an effective use of a very useful research tool.

REFERENCES

1. CARROLL, J. B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, **10**, 1–19.
2. CATTELL, R. B. Psychological measurement: ipsative, normative, and interactive. *Psychol. Rev.*, 1944, **51**, 292–303.
3. DUDEK, F. J. The dependence of factorial composition of aptitude tests upon population differences among pilot trainees. *Educ. psychol. Measmt.*, 1948, **8**, 613–633; 1949, **9**, 95–104.
4. GUILFORD, J. P. Creativity. *Amer. Psychol.*, 1950, **5**, 444–454.
5. GUILFORD, J. P., & LACEY, J. I. (Eds.). *Printed classification tests*. Army Air Forces Aviation Psychology Research Program Report No. 5. Washington, D. C.: Government Printing Office, 1947.
6. GUILFORD, J. P., & MICHAEL, W. B. Approaches to univocal factor scores. *Psychometrika*, 1948, **13**, 1–22.
7. MICHAEL, W. B. Factor analysis of tests and criteria: A comparative study of two AAF pilot populations. *Psychol. Monogr.*, 1949, **63**, (3), Whole no. 298.
8. THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univ. of Chicago Press, 1947.

Received June 2, 1951.

A NOTE ON CHRISTIE'S: "EXPERIMENTAL NAÏVETÉ AND EXPERIENTIAL NAÏVETÉ"

LEWIS BERNSTEIN

Veterans Administration Hospital, Denver, Colorado

A recently completed investigation at the University of Colorado Laboratories, the first of a series suggested by Benjamin (1), relates directly to the provocative issues raised by Christie (2) concerning the pre-experimental experience of rats. Although the results of this study are not yet ready for publication, it is felt that some of our preliminary findings should be made known at this time.

Our study was a test of Benjamin's (1) relationship-reinforcement hypothesis. More specifically, an attempt was made to establish empirically that a relationship between the experimenter and experimental animals does affect the course of learning. Special emphasis was given to the effect of an interrupted relationship upon the course of retention during a period of experimental extinction. Relationship, for purposes of this investigation, was defined in terms of the amount of handling of the experimental animals.

Fifty albino rats were used in this study. Each litter was separated according to sex, and split into three groups at the time of weaning (21 days after birth).[1]

Group *EH* (extra handling) was handled and petted by the experimenter for ten minutes per day per animal, from weaning through completion of the experiment. This handling was in addition to the handling required for experimental procedures.

Group *IH* (intermediate handling) was not handled until the animals were 50 days of age, at which time they were tamed by handling, for a period of 10 days. Thereafter, they were handled only as required by experimental procedures. The purpose of this handling was to reduce situational anxiety. No attempt was made to establish a relationship with these animals.

Group *NH* (no handling) was not handled at all. Special procedures were devised for introducing these animals into and removing them from the experimental apparatus without handling.

The experiment began when all animals were 60 days old. They were trained to go to the lighted side of a T-shaped discrimination box in a corrective situation, under hunger-food tension. As each animal reached the criterion of learning, it was assigned to a new subgroup, and 40 extinction trials followed. For half the animals in Group *EH*, the daily extra handling was continued throughout the extinction trials; and for the other half, handling was interrupted. Similarly, half the animals in Group *NH* continued to be unhandled while the other half received handling for the first time. Group *IH* continued to be handled for experimental purposes throughout the extinction trials.

---

[1] The relevant pre-experimental variables suggested by Christie (2, p. 333) will be reported in a future publication.

Our data for the original learning show that in terms of (a) number of trials required to master the habit, and (b) number of errors, Group *EH* was significantly superior to Groups *IH* and *NH;* and Group *IH* was significantly superior to Group *NH.*

Our extinction data show that: (a) Animals who were handled throughout both parts of the experiment made significantly fewer errors than animals who were unhandled throughout the experiment, or who received intermediate handling throughout the experiment. (b) Unhandled animals who were handled for the first time during extinction trials made fewer errors than animals who continued to be unhandled. This difference, which was not statistically significant, can possibly be explained by the fact that the handling of these animals lasted for only four days. (c) Animals who were handled throughout both parts of the experiment made significantly fewer errors than animals who were handled for the first time during extinction trials. (d) Handled animals with whom the relationship was interrupted made significantly more errors than animals with whom the relationship was continued during extinction trials. In fact, the animals with whom the relationship had been interrupted made more errors than the unhandled animals. In other words, an interrupted relationship produced more errors than a minimal relationship.

It was noted throughout the experiment that the extra-handled animals were more active and lively than the unhandled animals. This was apparent not only in maze behavior, but also in home cage activity. In the maze, the handled animals vigorously explored the apparatus —testing the wire-mesh covers, learning to open sliding doors with paws and nose, and pushing the glass food-dish aside. In marked contrast, the unhandled animals ran directly to the goal-box, without any apparent exploratory behavior. After eating the goal-food, they entered directly into the delay-chamber to await the next trial, without any exploration of the goal-box. None of the unhandled animals learned to open the doors in the apparatus. It was also noted that the unhandled animals attempted to hoard food following their daily feeding periods, whereas there was no such attempt on the part of the handled animals. On this basis, a repetition of Hunt's (4) hoarding study, introducing the variable of relationship, is being planned.

These quantitative, as well as qualitative, findings are offered as further evidence for Christie's point of view in regard to pre-experimental experience. However, these same findings suggest an alternate interpretation to Christie's (2) and Hebb's (3) emphasis on the importance of early exploratory training. Possibly a second factor was operative in Hebb's study: the pet animals may have had the benefit of a relationship with the experimenter as well as a richer experience. Of the two possible variables operative, Hebb has chosen to use the one of broader experience as an explanatory concept. It is of some interest, therefore, to note that Shurrager (5), whose animals were treated in a fashion similar to Hebb's, emphasizes the relationship between the animals and the experimenter, rather than the concept of broader experience.

Being aware of these two variables, our animals were handled just outside their living cages in order to eliminate, as far as possible, the contamination of the relationship variable with that of broader experience. For this same reason, all animals were housed in the experimental room and none was taken outside the room. Furthermore, the position of the cages on the rack was systematically shifted twice each week, so that every cage twice occupied each position on the rack. To compensate for any additional experience which might have accrued to the handled animals by being handled outside their cages, the unhandled animals were required to run through a portable alleyway, placed between the home cages and the feeding cages, in order to obtain their daily ration. The handled animals were placed in the feeding cages by hand.

We do not mean to minimize the importance of Hebb's and Christie's emphasis on the value of early exploratory experience. We do wish to suggest, however, that in light of our findings, the following hypothesis might be tested: with breadth of early exploratory experience held constant, the group of animals with whom the experimenter establishes a relationship (through extra petting and handling) will show learning superior to that of the animals that have only broad early exploratory experience.

Finally, in connection with Christie's (**2**) suggestion of investigating genetic hypotheses, using animal subjects, our data show that the extra-handled and unhandled animals were equated for weight at weaning. By the age of 46 days, the mean weight gain of the extra-handled animals was significantly greater than that of the unhandled animals. Unfortunately, it cannot be stated at the present time whether this was due to better physiological use of the food consumed or to greater quantities of food consumed. We hope to resolve this question in future experiments by weighing the animals' food.

REFERENCES

1. BENJAMIN, J. D. Methodological considerations in the validation and elaboration of psychoanalytical personality theory. *Amer. J. Orthopsychiat.*, 1950, **20**, 139–156.
2. CHRISTIE, R. Experimental naïveté and experiential naïveté. *Psychol. Bull.*, 1951, **48**, 327–339.
3. HEBB, D. O. *Organization of behavior*. New York: Wiley, 1949.
4. HUNT, J. McV. The effects of infant feeding frustration upon adult hoarding. *J. abnorm. soc. Psychol.*, 1941, **36**, 338–360.
5. SHURRAGER, P. S. "Spinal" cats walk. *Sci. Amer.*, 1950, **183**, 20–22.

Received August 8, 1951.

A NOTE ON THE PROBLEM OF HOMOGENEITY-HETEROGENEITY IN THE USE OF THE MATCHING METHOD IN PERSONALITY STUDIES

PAUL F. SECORD
Emory University

The matching method has been used in a good many personality studies and is particularly suitable for investigations utilizing data which cannot be broken down into quantitative scores or data which it is not desirable or feasible to analyze. Thus, an investigator may have judges match Rorschach test protocols with personality sketches in an attempt to determine the validity of the Rorschach test for the diagnosis of personality. One persistent problem encountered by the investigator using the matching method is that of the homogeneity or heterogeneity of the specimens being matched. The problem may be stated as follows:

> In order to make the task of the judges a reasonable one, it is necessary to divide the total number of specimens into smaller groups. Each judge then matches the test specimens in each of the smaller groups with items in the corresponding criterion group; for example, he may match several sets of five handwriting specimens with corresponding sets of five personality sketches. It is obvious that lack of similarity between the specimens within any one set will make the judgments easier and greater similarity will make them more difficult. The problem, then, is one of the degree of homogeneity within the total sample of specimens and within each of the subgroups, and the relationship of this homogeneity to the conclusions which may be drawn concerning results obtained.
>
> A study by Arnheim (1) frequently quoted with approval may be used to illustrate the fallacy of ignoring this problem. In this study, only three specimens (of handwriting) were used, those of Michelangelo, Raphael, and Leonardo da Vinci. These samples varied greatly, one being very bold, another very fine and delicate, and the other, intermediate. This heterogeneity of specimens makes matching with some other variable much easier than it would be if the specimens were similar. If the specimens representing the other variable are also heterogeneous, the task becomes still easier. An additional question here, of course, is the representativeness of the sample. With so few specimens at least one of them may correspond in character to the variable being matched with it, not because of any real relationship, but because of incidental selection of the sample. This could easily give statistically significant results when a large number of *judges* are used (2, p. 432).

The reader familiar with Brunswik's writings (2) or the note in this Journal by Hammond (3) will discern that the problem under discussion is a special case of the more general problem of representativeness of the "situation" in psychological experiments. Too often situation representativeness is slighted, and only population representativeness carefully considered in psychological research.

Vernon (3) in his excellent 1936 review of the matching method, identified homogeneity-heterogeneity as the most important problem in the use of this method, but his solution to it is compressed into a single sentence and apparently has been overlooked, since all of the studies employing the matching method which have come to the attention of the writer in more recent years have either ignored the problem or have not dealt with it in a completely adequate fashion.

The solution to the problem is to treat *specimens* for matching with the same sampling precautions that *persons* in the more usual experiment are treated. The total sample of specimens must be drawn from a defined population by strictly random methods, meeting the following well-known conditions: (*a*) each specimen in the population must have an equal chance of being included in the sample, and (*b*) the drawing of one specimen must in no way affect the drawing of another. The second crucial step is involved in the formation of the subgroups of specimens. This must be done by: (*a*) assigning random numbers to all of the specimens by use of a table, (*b*) arranging the specimens in consecutive order, and (*c*) choosing each subgroup by taking the specimens in order of their assigned number until each subgroup is completed.

Following the above procedure will give reasonable assurance that the conclusions drawn by the experimenter can be applied to the population from which the sample was drawn, since the degree of heterogeneity in the subgroups of the sample will approach that in the population in question, provided the sample is of a reasonable size (at least 40 or preferably more). Failure to follow this procedure strictly limits conclusions to the particular sample studied, and obviously such conclusions are of little value.

REFERENCES

1. ARNHEIM, R. Experimentell-psychologische Untersuchungen zum Ausdrucksproblem. *Psychol. Forsch.*, 1928, **11**, 1–132.
2. BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univ. of Calif. Press, 1947.
3. HAMMOND, K. R. Subject and object sampling—a note. *Psychol. Bull.*, 1948, **45**, 530–533.
4. SECORD, P. F. Studies of the relationship of handwriting to personality. *J. Personality*, 1949, **17**, 430–448.
5. VERNON, P. E. The matching method applied to investigations of personality. *Psychol. Bull.*, 1936, **33**, 149–177.

Received June 18, 1951.

TESTS OF HYPOTHESES: ONE-SIDED *VS.* TWO-SIDED ALTERNATIVES[1]

LYLE V. JONES
University of Chicago

Psychological literature abounds with experimental studies which utilize statistical tests of the significance of differences between two groups of subjects. Most of these studies present tests based upon either the distribution of Student's t or upon the distribution of χ^2. Since the comparison of an experimental group with a control group of subjects is so fundamental to the experimental method, and since statistical tests of significance are appropriate for testing hypotheses regarding differences between two groups of subjects, it would seem important to correct a common misconception concerning the application of these tests of hypotheses.

One model for a test of significance of mean difference, the more familiar model, is that in which we test the null hypothesis, H_0, against a set of two-sided alternatives, H_1. We might formalize this test,

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0,$$

where μ_1 is the mean of the population represented by one sample and μ_2 is the mean of the population represented by a second sample. Assuming scores X_1, from the first population, and scores X_2, from the second, both to be distributed normally, and assuming the population standard deviations to be equal, we may find

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}},$$

where

$$s = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \overline{X}_1)^2 + \sum_{j=1}^{N_2} (X_{2j} - \overline{X}_2)^2}{N_1 + N_2 - 2}}$$

and N_1 and N_2 are the numbers of individuals in the samples from the first and second populations.[2] Having stipulated a desired confidence level, α, we may enter the t table (1) with N_1+N_2-2 degrees of freedom and a p-value equal to α to find a critical value of t, t_c. If the absolute value of the observed t exceeds t_c, we reject H_0 in favor of H_1; otherwise, we accept H_0. In Figure 1 appears a distribution of t showing, graphically, the nature of this decision. This distribution corresponds to the sampling distribution of mean differences, under the null hypothesis. For any value of t to the right of t_c or to the left of $-t_c$, we reject H_0. The two shaded tails of the distribution, taken together, make up α per cent of the total area under the curve.

---

[1] This note was prepared while the writer was a National Research Council Fellow.

[2] Of course, if the two samples are not independently selected, we should make use of the correlation between them in the determination of t.

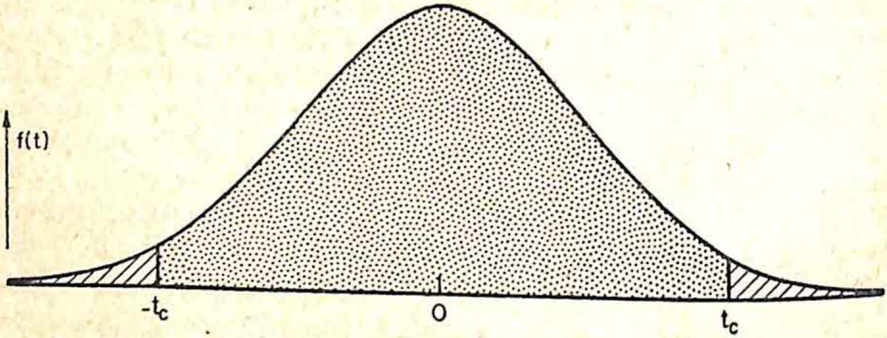


FIG. 1. THE TWO-TAILED TEST MODEL.

The model above, the test of the null hypothesis against two-sided alternatives, is the one used most often by investigators in psychology. Yet in many cases, probably in most cases, it is not the test most appropriate for their experimental problems. More often than not, in psychological research, our hypotheses have a *directional* character. We are interested in whether or not a given diet *improves* maze performance in the rat. We hypothesize that the showing of a particular motion picture to a group of individuals would lead to a *more tolerant* attitude toward certain racial minorities. We wish to test whether or not anxious subjects will respond *more actively* than normal subjects to environmental changes which might be perceived as threatening. In each case, theoretical considerations allow the postulation of the direction of experimental effects. The appropriate experimental test is one which takes this into account, a test of the null hypothesis against a one-sided alternative.

In the one-sided case we test H_0 against H_1, where

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 > 0.$$

Under the identical assumptions of the two-sided model we may calculate t as before. Again a confidence level, α, is stipulated. The distinction between the one-sided test and the two-sided test arises in the determination of the critical value, t_c. In the present case this critical value is found by entering the t table with N_1+N_2-2 degrees of freedom, as before, but with a p-value equal to 2α. If our observed t is greater than this t_c we reject H_0 in favor of H_1; if t is less than t_c we accept H_0. The t distribution in Figure 2 exemplifies this procedure. A value of t to the right of t_c leads to the rejection of H_0, the acceptance of H_1. While the shaded area under the curve once again represents α per cent of the total area, the shaded portion is restricted, in this case, to one tail of the distribution.

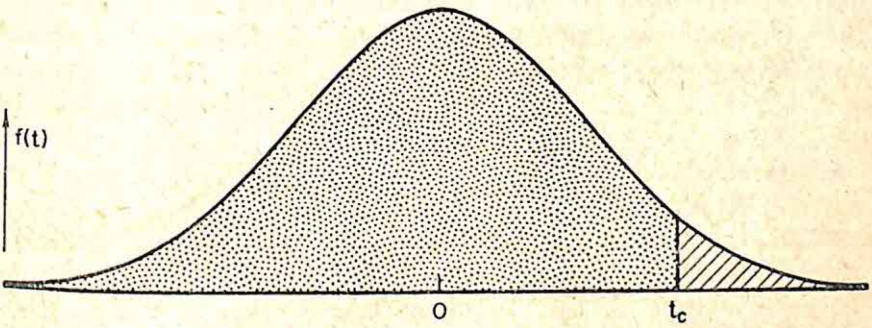


FIG. 2. THE ONE-TAILED TEST MODEL.

It might be noted that with this formulation of the one-tailed test there is no allowance for the possibility that the true difference, $\mu_1-\mu_2$, is negative. In the type of problem for which the one-tailed test is suited, such a negative mean difference is no more interesting than a zero difference. In fact, the hypotheses for the one-sided case might be

$$H_0\colon \mu_1 - \mu_2 \leqq 0$$

$$H_1\colon \mu_1 - \mu_2 > 0.$$

In order to determine a sampling distribution under H_0 we should consider the "worst" of the infinite alternatives under H_0, i.e., that alternative which would make the decision between H_0 and H_1 a most difficult one. Clearly, the decision would be more difficult if the true mean difference were zero than if the true difference were any negative value. Hence we would proceed exactly as in the preceding one-tailed case, utilizing, for our test, the distribution of t based upon the same sampling distribution of mean differences as before. The confidence level should be doubled to provide the p-value for entering a table to find a critical t_c, or, if it is desired to ascertain the p-value corresponding to an observed t, the correct value is one-half that given in the typical table of t.

While the one-tailed test has been exemplified here as a test of mean difference, based upon the t distribution, it is limited in application neither to mean difference problems nor to the t statistic. Indeed, whereever an alternative to the null hypothesis is stated in terms of the direction of expected results, the one-tailed test is applicable.

The failure, among psychologists, to utilize the one-tailed statistical test, where it is appropriate, very likely is due to the propagation of the two-tailed model by writers of text books in psychological statistics. It is typical, in such texts, to find little or no attention given to one-tailed tests.[3] Since the test of the null hypothesis against a one-sided alternative is the most powerful test for all directional hypotheses, it is strongly recommended that the one-tailed model be adopted wherever its use is appropriate.

REFERENCE

1. Fisher, R. A. & Yates, F. *Statistical tables for biological, medical and agricultural research.* Edinburgh: Oliver & Boyd, Ltd., 1938.

Received August 8, 1951.

---

[3] One notable exception occurs in A. L. Edwards' *Experimental Design in Psychological Research*, where the two cases are clearly differentiated.

SOME COMMENTS ON THISTLETHWAITE'S PERCEPTION OF LATENT LEARNING

HOWARD H. KENDLER
New York University

Controversy frequently has an invigorating effect upon research and a confusing effect upon theory. The area of "latent learning" is an example; with each successive experiment the theoretical picture appears to become more cloudy. The time would seem to be ripe for a clarifying review of the issues. Unfortunately Thistlethwaite's attempt (14) is unsuccessful because he ignores certain basic issues. I shall avoid making a point by point rebuttal to his thirty-page review, but shall instead limit my comments to three major issues.

1. It appears flagrantly inconsistent to admit, as do Thistlethwaite and other cognitive theorists, that some reward was present during the "no food" trials of the California studies while simultaneously interpreting the results to indicate that reward is not necessary for learning. If some reward did exist, as is indicated by the reduction in errors of the *S*s during the latent learning period (2, 16), as well as by the evidence of other studies (7, 10, 11), then the point at issue is not whether learning requires reward but rather how different *amounts* or *types* of reward affect learning. Admittedly the concept of reinforcement is vague at the low end of the intensity dimension. I suspect that this vagueness will disappear as the motivational component of *S-R* reinforcement theory continues to be developed. One must recognize, however, that it would be foolish to dispense with the concept of reinforcement because of its lack of precision at certain intensities. The concept has been too helpful in integrating large amounts of data to be discarded. Meanwhile it would be unfortunate to confuse the issue of whether learning requires reward with the issue of the relationship between learning and different amounts or types of reinforcement.

2. The underlying theme of Thistlethwaite's review of latent learning data appears to be that those studies which "demonstrate" latent learning are sound, well-designed experiments while those that fail to provide positive evidence are ill-conceived and inappropriate to cognitive theory. His analysis of the Kendler and Mencher study (9) is typical of his treatment of studies which report evidence embarrassing to cognitive theory.

1. Thistlethwaite initially criticizes the Kendler and Mencher study by suggesting that the failure to obtain latent learning could be attributed to the pronounced preference the *S*s had for the black alley as compared to the white alley. No mention is made, however, of Grice's failure (5) to obtain latent learning in a similar situation with the entire T-maze painted a flat

black. Nor is any mention made of the fact that Kendler (**8**) reported learning of the position of food and water in a T-maze having one side painted black while the other was left unpainted. In this study, the *S*s were permitted to consume the goal objects.

2. The second criticism leveled at the Kendler and Mencher study was that it used a procedure in which 50 per cent of the trials were forced. Again, no mention was made of the fact that Kendler (**8**) was able to obtain learning of the position of food and water in spite of his use of a 50 per cent forced trial procedure. Likewise ignored were other studies (**4, 15**) which demonstrated that cognitions (or habits) were established during forced trials when rewards were operating. Nor did Thistlethwaite bother to relate the results of Walker, Knotter, and De Valois (**17**), which he had interpreted as being favorable to cognitive theory, to his second criticism. These experimenters found no difference in the results of those *S*s who had 50 per cent of their trials forced as compared with the group that was permitted to choose freely on all training trials.

3. The third and final criticism offered by Thistlethwaite was that " . . . the technique used by Kendler and Mencher of comparing *over the total 20 test trials* the performance of a group which found food on the same side as during the original thirst training with that of a group which found the food reversed to the opposite side, presupposed that learning is a gradual process of strengthening habits. On the basis of . . . non-continuity interpretations of learning . . . , one might predict that the reversed group would rather quickly give up a maladaptive hypothesis and learn within a very few trials to go to the now-correct side. The comparison of total performance scores over a long relearning period might well obscure, rather than enhance, any initial superiority of the constant group." Such a statement is misleading because it fails to mention that other response measures were reported in the Kendler and Mencher study. The criticism would not be relevant to the first test trial, the results of which fail to provide any evidence of latent learning. Nor did the analysis of the records of individual *S*s provide any positive evidence. Furthermore, it seems questionable to raise such a criticism since the non-continuity interpretation of learning has been so badly discredited in recent years (**3, 6**). Even such cognitively oriented psychologists as Ritchie (**12**) and Bitterman (**1**) have reported data in opposition to such an interpretation. If the assumptions underlying a criticism are invalid, there is no point in offering such a criticism.

Thistlethwaite's analysis of the Kendler and Mencher study, which is typical of his treatment of findings embarrassing to cognitive theory, is neither "cricket" nor "science" for several reasons. First, of course, he fails to consider all the relevant data. It might be argued that latent learning can occur only when there is a subtle combination of several variables. The point can be a reasonable one when the necessary combination of variables is specified beforehand. Otherwise, discordant results can be attributed, as Thistlethwaite has done, to anything and everything. Second, he has ignored the fact that experiments can be designed only in terms of the *explicit* requirements of a theory. The "negative findings" of Kendler and Mencher, and others, were obtained in experimental situations which met all the *stated specifications* of cognitive theory. If the results were negative the blame belongs to the theory. Finally, in Thistlethwaite's eagerness to perceive studies as being favorable to cognitive theory, he practically ignores the problem of why the "positive findings" of irrelevant-incentive learning have revealed such a small degree of learning. The problems, it seems to me, are for the cognitive theorist to demonstrate why such learning is so poor and for the *S-R* reinforcement theorists to show why any learning takes place. It is interesting to note the differences between Thistlethwaite's attempt to answer the first question and Spence's attempt (13) to answer the second question.

3. Finally, I would like to raise a logical point which has evaded other cognitive theorists as well as Thistlethwaite. One of the major conclusions drawn from the results of the initial California latent learning studies (2, 16) was that maze performance did not necessarily mirror maze learning. *S-R* psychologists have no argument with such a conclusion. A reasonable objection, however, might be raised because of the cognitive theorists' failure to apply this principle consistently to all aspects of their latent learning experiments. For example, Tolman and Honzik (16) have indicated that the difference between the performance of the "food" and "no food" groups during the "latent learning period" *did not reflect differences in learning*. Their reasoning was based upon the fact that both groups *performed at the same level of efficiency* following the introduction of food. If different levels of performance, however, need not imply different levels of learning, then according to the basic assumption underlying their analysis (maze performance does not necessarily mirror maze learning), it is possible that *the same level of performance need not imply the same level of learning*. That is, if one assumes, as did Tolman, an absence of an isomorphic relationship between learning and performance, it is possible that under certain conditions different degrees of learning can produce the same level of performance. The cognitive theorists have never considered this latter point but have passively accepted the equality of performance of the "food" and "no food" groups on the test trials of the latent learning experiments as evidence of equal learning.

If the cognitive theorists felt compelled to build their theoretical system around the California latent learning studies, it would have seemed advisable to test further the assumption that the similar performance of the "food" and "no food" group really indicated equal learning. An obvious experimental design would be one involving the retroactive inhibition paradigm, i.e., some activity interpolated between the training and test trials. If the results of such an experiment demonstrated that both groups resisted the effects of an interpolated activity equally well, then greater confidence could be placed in the conclusion of the cognitive theorists that the two groups had equal learning in spite of their different motivational-reward conditions. If, however, the performance of the two groups differed after the interpolated activity, then the interpretation of the cognitive theorists that there was equal learning during the "no food" period would be suspect. Kanner,[1] in a recently completed experiment, similar in design to the Tolman and Honzik study, found that when reward was introduced after an interpolated activity, the performance of the "food" group was significantly superior to that of the "no food" group. Such data would be strong presumptive evidence that the *S*s in the California latent learning studies, as well as in Kanner's preliminary study in which he was able to obtain latent learning, did *not* have equal learning; for if they had, their performance following the interpolated activity would not have differed.[2]

REFERENCES

1. BITTERMAN, M. E., & COATE, W. B. Some new experiments of the nature of discrimination learning in the rat. *J. comp. physiol. Psychol.* 1950, **43**, 198–210.
2. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113–134.
3. EHRENFREUND, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol. Psychol.* 1948, **41**, 408–422.
4. GRICE, G. R. An experimental study of the gradient of reinforcement in maze learning. *J. exp. Psychol.*, 1942, **30**, 475–489.
5. GRICE, G. R. An experimental test of the expectation theory of learning. *J. comp. physiol. Psychol.*, 1948, **41**, 137–143.
6. GRICE, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, **38**, 633–642.
7. KARN, H. K., & PORTER, J. M., JR. The effects of certain pretraining procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.
8. KENDLER, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 461–469.
9. KENDLER, H. H., & MENCHER, HELEN C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, **38**, 82–88.
10. MEEHL, P. E., & MACCORQUODALE, K. A failure to find the Blodgett effect, and some secondary observations on drive conditioning. *J. comp. physiol. Psychol.*, 1951, **44**, 178–183.

---

[1] KANNER, J. H. The influence of reward in the California latent learning studies. (In preparation.)

[2] An attempt to explain why the "food" and "no food" groups performed equally well following the introduction of food in the California studies (**2**, **16**) as well as in Kanner's preliminary study, will be presented in Kanner's paper. It is based upon the assumption that the motivational-reward conditions of the two groups were drastically different. This assumption is, of course, quite different from the one tacitly accepted by most psychologists, namely, that both groups had the same motivation except that one group had an additional source of reward (food). It should be recognized, however, that this problem is not relevant to the present discussion. The major point of my analysis has been to question whether such "equal performance" implied "equal learning" and the results of Kanner's studies indicate that it did not.

11. Reynolds, B. A repetition of the Blodgett experiment on "latent learning." *J. exp. Psychol.*, 1945, 35, 504–516.
12. Ritchie, B. F., Ebeling, E., & Roth, W. Evidence for continuity in the discrimination of vertical and horizontal patterns. *J. comp. physiol. Psychol.*, 1950, 43, 168–180.
13. Spence, K. W., Bergmann, G., & Lippitt, R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.*, 1950, 40, 539–551.
14. Thistlethwaite, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97–129.
15. Thompson, M. E. An experimental investigation of the gradient of reinforcement in maze learning. *J. exp. Psychol.*, 1944, 34, 390–403.
16. Tolman, E. C., & Honzik, H. C. Introduction and removal of reward, and maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, 257–275.
17. Walker, E. L., Knotter, Margaret C., & DeValois, R. L. Drive specificity and learning: The acquisition of a spatial response to food under conditions of water deprivation and food satiation. *J. exp. Psychol.*, 1950, 40, 161–168.

Received June 14, 1951.

THE BLODGETT AND HANEY TYPES OF LATENT LEARNING EXPERIMENT: REPLY TO THISTLETHWAITE

IRVING MALTZMAN
University of California at Los Angeles

This paper will be restricted to an analysis of the Blodgett and Haney types of latent learning experiment and to Thistlethwaite's comments upon them. Its purpose will be to show that these studies, designated types I and II by Thistlethwaite, do not demonstrate latent learning, and that their results can be interpreted within the framework of Hull and Spence's reinforcement theory of behavior.

According to Thistlethwaite, "In experiments of the Blodgett type, the question of whether there are *some* manifestations of *some* learnings during the latent period is comparatively unimportant. It is not necessary that there be no evidence of a tendency toward error reduction during the unrewarded (or slightly rewarded) trials in order for the unmanifested learning to qualify as 'latent learning.' The question of whether the abrupt increase in efficiency of running the maze upon the introduction of food is significantly greater than the comparable improvements of a control group is a statistical question. If it is in any given experiment, then that experiment may be interpreted as yielding positive evidence of latent learning" (15, p. 100).

In his review of the individual experiments Thistlethwaite does not refer again to manifestations of learning during the latent period of experiments that presumably demonstrate latent learning, viz, those of Blodgett (1), Tolman and Honzik (16), Herb (6), Haney (5), and Daub (3). In most of these experiments a large amount of improvement occurred during the latent period. As the subsequent analysis will attempt to show, such improvement has an important bearing on the interpretation of the results of these experiments.

HAS "LATENT LEARNING" BEEN EXPERIMENTALLY DEMONSTRATED

1. *The Blodgett Design.* Blodgett's (1) comparison of the drop in errors between the control group (C) and the three day experimental group (X_3) gave a critical ratio of 1.338. The comparison of groups C and X_7 gave a critical ratio of 1.544. Blodgett stated that the obtained differences were significant at approximately the 6 and 9 per cent levels respectively (1, p. 123). It is apparent that only one-half of the normal probability distribution function was used to arrive at these confidence levels. For both tails of the distribution critical ratios of the size ob-

tained are significant at the 12 and 18 per cent levels respectively. Blodgett did not state that he was using a single-tailed hypothesis, and he did not explicitly predict the direction of the results. The use of a single-tailed hypothesis under these circumstances may be seriously questioned. If this is the case, a significant amount of latent learning was not obtained with the criterion used.

Thistlethwaite does not present the reliability of the results, or lack of them. He does remark that Tolman and Honzik (16) obtained more significant results and that evaluation by another criterion yielded significant results (15, p. 102). He also fails to mention the results of Blodgett's experiment in terms of the time criterion, results which failed to show latent learning.

In addition to a comparison of the size of the reduction of errors or running time Blodgett employed a second criterion of latent learning. The latter method compared the drops in errors following reward of the experimental groups with interpolated drops in the control group—from *performance* levels corresponding to those from which the drops in the experimental groups began.

Blodgett reported the confidence levels as approximately the 7 and 2 per cent levels respectively. These, again, are incorrectly based on a single-tailed hypothesis. The results of this procedure are, of course, of greater statistical significance than those obtained with the first criterion used. But they do not constitute a demonstration of latent learning, at least as far as Hull's theory of behavior is concerned. The performance curves of groups X_3 and X_7 show some reduction of errors during the latent period. For X_3 the error reduction on day three is equivalent to the effects of approximately 1.6 reinforcements in group C. Performance of X_7 on the day reward was introduced was equivalent to performance in group C after approximately 2.3 reinforcements.

Hull (9) assumes that habit strength grows as a negatively accelerated exponential function of the number of reinforcements. This implies that successive increments in habit strength become progressively smaller. It follows, therefore, that the increment in habit strength after one rewarded trial in the experimental groups would be greater than the increment in habit strength after 1.6 or 2.3 reinforcements in the control group. Latent learning in terms of this second criterion can thus be readily accounted for in terms of Hull's reinforcement theory of behavior.

A third possible criterion of this type of latent learning is the saving score obtained by comparing the difference between the differences in performance of the control and experimental groups on the day reward is introduced and on some subsequent day. This scoring method would take into account initial differences in performance among the different groups, a factor that the other methods do not reflect. In terms of absolute and relative saving scores, performance of Blodgett's group C was superior to that of groups X_3 and X_7. Whether or not the saving scores are significantly different cannot be determined. However, the direction and size of such scores in all experiments of this kind indicate that they probably are significant. If latent learning had developed, and if improvement during the unrewarded period is of no importance as Thistlethwaite states (**15**, p. 100), there should be no difference in saving scores between the control and experimental groups.

Blodgett's experiment was repeated by Reynolds (**12**) who used the same six-unit T-maze pattern. In Reynolds' first study maze-wise rats were run as a control group (*C*) and as a seven-day experimental group (X_7). Since the experimental *S*s learned the maze prior to the introduction of reward, latent learning was not exhibited. Reynolds was criticized for using maze-wise rats (**2**, **15**). But the fact that he also ran a control group and seven- and three-day experimental groups of naive *S*s has been minimized by Blodgett (**2**), Hilgard (**8**), and Thistlethwaite (**15**).

In the second study the naive *S*s of X_7 again learned prior to the introduction of reward producing fewer errors than the X_7 maze-wise rats. Despite this, Blodgett (**2**) and Thistlethwaite (**15**, p. 103) have argued that the maze-wise rats probably learned so rapidly because of externalization of drive.

Contrary to the impression created by Thistlethwaite and other critics of reinforcement theory (**2**, **8**) the naive X_3 *S*s did not learn prior to the introduction of reward. They committed 3.03 mean errors on day three compared to 2.72 mean errors of Blodgett's X_3 group on the same day. Nevertheless, these *S*s did not show latent learning according to Blodgett's criterion. The largest drop in errors of group *C* was larger than the drop of X_3 following the introduction of reward. The same results were obtained in terms of running time.

The fact that the naive *S*s of X_3 did not learn prior to the introduction of reward and did not show latent learning is not mentioned by Thistlethwaite. It is only indirectly implied by the criticism that repetitions of Blodgett's experiment do not get the same results because the reward used was not as palatable or the *S*s were not as highly motivated as in Blodgett's study (**15**, p. 103). It is difficult to see the relevance of these criticisms, since the absolute decrease in errors or time of the experimental groups is not the criterion of latent learning. The difference in the size of the drops for the control and the experimental groups is the commonly cited criterion. Differences in palatability or strength of drive should affect the performance of control and experimental groups equally. Furthermore, it is by no means clear that all the experiments of this type, apparently showing latent learning, used a palatable mash instead of hard dog chow, or that the *S*s were both hungry and slightly thirsty or only hungry.

Although Blodgett's experiment has been criticized as failing to demonstrate latent learning, this by no means constitutes an adequate criticism of all latent learning experiments of this type. Thistlethwaite states that other experiments give more striking evidence of latent learning. "Blodgett used a six-unit T-maze. Tolman and Honzik used a 14-unit T-maze and obtained, following the introduction of reward, sudden drops in error and time scores which were steeper and of greater statistical significance than Blodgett's" (**15**, p. 102).

This statement is rather surprising because, contrary to the impression created, Tolman and Honzik did not determine the statistical significance of the drops in errors and time following the introduction of food in the experimental group as compared with the largest drop in the control group. They compared mean error and time increases of the HR and HR-NR groups, and mean drops in errors and time between HNR and HNR-R groups (**16**, pp. 164, 165). The first comparison indicated the effects of withdrawal of reward while the second indicated the effects of the introduction of reward as contrasted with unrewarded performance. A comparison of this sort was not made between HNR-R and HR. They did compare the performance of the two groups from the twelfth day on, following the introduction of reward on the eleventh day. In terms of errors the HNR-R group was significantly superior at the 2 per cent level. There was no significant difference in terms of time scores.

However, prior to the introduction of reward the experimental group (HNR-R) showed an amount of improvement equal to that of the control group (HR) after seven trials. If the improvement occurring prior to the introduction of food reward is of no importance as nonreinforcement theorists insist, saving scores between days 1 and 17 of the control group and days 11 and 17 of the experimental group should not differ. The control group showed greater improvement indicating that when the level of performance at the point of introduction of reward is taken into account, it cannot be concluded that latent learning is demonstrated by Tolman and Honzik.

2. *The Haney Design.* A second experimental design used to investigate latent learning was employed by Haney (**5**). An experimental group of hungry rats was permitted to explore a 14-unit T-maze for four days, 18 hours per day. The control group of *S*s spent the same amount of time in a rectangular maze. Neither group was rewarded with food in the maze during this period. Both groups were then run hungry and rewarded one trial per day in the T-maze for 18 days. The experimental *S*s that were permitted to explore the T-maze thus had an opportunity to acquire knowledge about the maze which the control *S*s run in the rectangular maze could not, according to the nonreinforcement theorist.

In terms of the mean errors committed over the 18-day period the experimental group was greatly superior to the control group. A critical ratio of 8.8 was obtained. Running times for the two groups were not significantly different.

According to the sign-gestalt interpretation the experimental *S*s possessed knowledge of the sequence of paths on the first test trial, but no knowledge of the presence of food. The knowledge of "what leads to what" would therefore continue to be latent on the first trial. The experimental *S*s' performance on this trial should be no different from that of the control *S*s that did not have knowledge of either the sequence of paths or the presence of food. However, on the first day's run prior to commerce with the goal object the experimental group committed half as many errors as the control group (**5**, p. 327). Haney's experimental group showed approximately the same per cent of error reduction as Reynolds' seven-day experimental group prior to the introduction of reward. Thistlethwaite insists, and correctly so, that Reynolds' seven-day group did not meet the conditions necessary for a test of the latent learning hypothesis, because they had learned prior to the introduction of reward. The same conclusion must hold for Haney's experiment. The experimental group already learned the maze to a great extent prior to the introduction of reward.

Furthermore, the drop in errors between trials one and two for the control group was greater than for the experimental group. Latent learning was not obtained according to this criterion. Nevertheless Thistlethwaite states that latent learning was demonstrated in this experiment (**15**, p. 103).

Daub (**3**) repeated Haney's experiment in order to determine the effects that doors which prevent retracing have on performance in the test trials run in the T-maze. The criticisms of Haney's experiment apply as well to Daub's study. The control and experimental groups were not at the same level of performance at the time of introduction of reward, and a control group showed the largest decrease in errors between successive trials, although it did not occur between the first and second trial. In both studies, saving scores indicate a greater amount of improvement by the control than experimental groups following the introduction of food reward.

Another variation of the introduction of reward type of latent learning experiment has been conducted by Herb (**6**). A 14-unit T-maze was used in this study. Prior to training, *S*s in the control and experimental groups were given six days of preliminary training totaling 30 trials in not more than two units of the maze. The control group of hungry *S*s was then given one trial per day for 17 days in which food was found at the end of each blind alley. The experimental group received one trial per day for 10 days while hungry, but received no food in the maze. From days 11 to 17 food was found in each of the blind alleys entered.

The experimental group showed a large reduction in errors between the first and the tenth days in the absence of food reward. The difference was significant at the one per cent level of confidence. Following the introduction of reward in the blind alleys there was a large increase in the number of errors. Herb concluded that latent learning was evidenced by the significantly greater increase in errors committed by the experimental group on the two trials following the introduction of reward as compared with the largest increase in errors of the control group over a two-day period. It should be pointed out that a smaller difference between the two groups would be obtained if only one day following the introduction of reward in the experimental condition was compared with the largest drop in the control condition over a period of one day. Once again, here is the arbitrary selection of a criterion on the grounds that it will provide the most convincing evidence of latent learning.

The control group made a total of 145 errors or blind alley entrances on the first training day. On the first training day that the experimental group received food, day eleven, they committed 102 errors. The rate of acquisition according to Hull's theory should therefore differ in the two groups as a result of this initial difference in performance.

ALTERNATIVE INTERPRETATION OF RESULTS SECURED IN "LATENT LEARNING" EXPERIMENTS

In every study of the effects of the introduction of reward that was discussed, except Reynolds', latent learning was said to have been demonstrated by Thistlethwaite and the respective authors. However, it was shown that in some of these studies (1, 5) latent learning did not even occur according to its commonly accepted criterion. Saving scores indicated that none of the experiments in which this criterion could be applied clearly demonstrated latent learning. In all of these studies the performance levels of the experimental and control groups were different at the time food reward was introduced.

According to Hull's theory of behavior these different levels of performance reflect different absolute amounts of effective reaction potential (9). Since the negatively accelerated growth of reaction potential approaches an asymptote and the size of a given increment is a constant fraction of the difference between the asymptotic level and the absolute amount already acquired, the size of subsequent increments will differ for the two groups.

The different initial levels of performance, it is believed, are due primarily to the different amounts of habituation given the control and experimental groups in the training maze. In the Tolman and

Honzik experiment, for example, the experimental group received eleven more trials than the control group before the introduction of reward. Herb's experimental group had ten more trials than the control group in which to become adapted to the maze environment before food was found in the blind alleys.

This greater period of habituation permits a more complete extinction of responses incompatible with the correct responses of the training series. Estes (4) has recently shown that competing responses produce a depression in the initial phase of the performance curve reflecting the acquisition of strength of the correct response. These competing responses are probably reactions induced by fear of the new stimulus situation. Higginson (7) has shown that the arousal of fear produces an initial phase of positive acceleration in performance curves comparable to that displayed by the control *S*s in latent learning experiments of the kind discussed here. Karn and Porter (10) have also indicated the importance of habituation in the determination of latent learning. Improvement in *performance* during the pre-reward period in terms of a reduction in running time or errors is not due solely to *learning* or to the acquisition of habit strength, but to the elimination of the dominant competing responses as well. The additional days of habituation received by the experimental groups afford greater opportunity for the extinction of these responses prior to the introduction of reward.

In the Tolman and Honzik study the 17 food-rewarded trials of the control group, which may also be considered habituation trials, are less effective than the 11 habituation trials and only 5 rewarded trials of the experimental group. According to the present analysis this may occur because the increments in effective reaction potential following rewarded trials are different in the two groups.

For the control group in this type of latent learning study the largest increments in habit strength which develop during the early trials must compete with the reaction potential of incompatible responses. The effective reaction potential determining the arousal of the correct responses will thereby be reduced. This would produce a positively accelerated initial phase in the performance curve. When competing responses are finally extinguished and the reaction potential determining the correct response is dominant, the size of successive increments in habit strength will be relatively small. Furthermore, competing responses will be extinguished relatively slowly because these incorrect responses will acquire some additional habit strength as a result of generalization. This is not the case with respect to the growth of effective reaction potential in the experimental group. When reward is introduced after 11 unrewarded trials, competing emotional responses are to a great extent extinguished. As a result, the effective reaction potential determining the correct responses will not be greatly reduced by competition with the reaction potentials of incompatible responses. Therefore, the initial increments in effective reaction potential of the correct responses in the experimental group will be larger than any increment for the control group, and the rate of acquisition will be greater than that of the control group.

It is believed that the above analysis provides an account of the sudden improvement in performance following the introduction of reward in the type I and II latent learning experiments which is in accord with a reinforcement theory of behavior. It is clear, of course, that if there are adequate grounds, the account presented may be rejected or alternative hypotheses formulated without prejudicing the implications of saving scores or the reinterpretation of statistical results previously discussed.

This review has been restricted to the Blodgett and Haney type of introduction of reward experiment. The writer has attempted to show that Thistlethwaite and others have uncritically accepted the results of these experiments. The omission of any discussion of other kinds of latent learning experiment does not signify agreement with Thistlethwaite's review of these studies. Their discussion was omitted for lack of space, and because it was felt that most of them have been discussed adequately elsewhere (**11, 13, 14**).

REFERENCES

1. Blodgett, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113–134.
2. Blodgett, H. C. Reynolds' repetition of Blodgett's experiment on latent learning. *J. exp. Psychol.*, 1946, **36**, 184–186.
3. Daub, C. T. The effect of doors on latent learning. *J. comp. Psychol.*, 1933, **15**, 49–58.
4. Estes, W. K. Effects of competing reactions on the conditioning curve for bar pressing. *J. exp. Psychol.*, 1950, **40**, 200–205.
5. Haney, G. W. The effect of familiarity on maze performance of albino rats. *Univ. Calif. Publ. Psychol.*, 1931, **4**, 319–333.
6. Herb, F. H. Latent learning-non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, **29**, 247–256.
7. Higginson, G. D. The after-effects of certain emotional situations upon maze learning among white rats. *J. comp. Psychol.*, 1930, **10**, 355–373.
8. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century, 1948.
9. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
10. Karn, H. W., & Porter, J. M. The effect of certain pre-training procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.
11. Maltzman, I. M. An interpretation of learning under an irrelevant need.

Psychol. Rev., 1950, 57, 181–187.

12. Reynolds, B. A repetition of the Blodgett experiment on latent learning. *J. exp. Psychol.*, 1945, 35, 504–516.
13. Spence, K. W., & Kendler, H. H. The speculations of Leeper with respect to the Iowa tests of the sign-gestalt theory. *J. exp. Psychol.*, 1948, 38, 106–109.
14. Spence, K. W., Bergmann, G., & Lippitt, R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.*, 1950, 40, 539–551.
15. Thistlethwaite, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97–129.
16. Tolman, E. C., & Honzik, E. C. Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, 257–275.

Received April 30, 1951.

REPLY TO KENDLER AND MALTZMAN

DONALD THISTLETHWAITE
University of Illinois

Bertrand Russell once optimistically commented: "In science there is a technique of persuasion so effective that controversies seldom last very long. The technique consists of an appeal to evidence, not to emotions." No doubt it must seem strange to observers, both within and outside of psychology, that controversy over the Law of Effect (and more recently over its offspring, Reinforcement Doctrine) should persist so long. Part of the difficulty seems to lie in the fact that the participants in this controversy are still burdened with a number of pseudo problems. A more rigorous elimination of untestable hypotheses and continued appeal to evidence would seem to be the most promising means of bringing the controversy to an end. The critical comments of Kendler (**10**) and Maltzman (**13**) on my review (**20**) provide an opportunity to clarify the evidence, and possibly to adjust our differences of opinion concerning it.

REPLY TO KENDLER

Although Kendler says his remarks are limited to "three major issues" allegedly ignored by me, he does not bother to state any of these issues. Since the basic issues Kendler has in mind are conjectural, I will limit my remarks to the specific criticisms elaborated in his paper.

1. My conclusion that some of the experiments reviewed "constitute exceptions to the alleged indispensability of reinforcement in the learning process" was derived from a consideration of the inability of reinforcement theorists to account for all of the "negative instances" accumulated (pp. 121–123). Kendler's suggestion that this conclusion is based merely upon the results of the California studies is misleading.

The charge that I was "flagrantly inconsistent" in admitting that some error-reduction occurred during the latent period of acquisition in the Blodgett kind of experiment (type I) is unjustified. It is curious that Kendler ignores my discussion of this point. I argued that before we may grant the indispensability of reinforcement for learning *we must be able to demonstrate for each instance of learning* (*a*) *that some source of reinforcement was operative in the experimental setup, and* (*b*) *that the changes in responses which are taken as evidence of the learning can be deduced on the basis of this alleged reinforcement.* Let us briefly consider the first condition. The fact that error-reduction occurred during the latent period of some of the Blodgett-type experiments indicates the operation of some motivation producing progress through the maze during the unrewarded trials of these experiments. Kendler's conclusion from this observation is that the Blodgett type of experiment is irrelevant to the question of whether learning requires reward. His point is similar to one recently made by Spence (**19**). In discussing some of the irrelevant-incentive experiments (types III and IV) Spence argues that " . . . they have no bearing on the question of whether or not reinforcement is necessary for the hypothetical learning change to occur. All these experiments involve some form of motivation and a reinforcing state of affairs of some kind . . . . Thus the question of whether or not cognitive structures can be established . . . *without a reinforcing state of affairs of any kind* cannot be answered by this experimental design" (**19**, p. 277). Several points need to be clarified in this argument. First of all, it is not by any means clear that *all* of the latent learning and irrelevant-incentive learning experiments involved some form of a reinforcing state of affairs. It is certainly not true that all of these experiments provided primary need-reduction contiguous with the response to be learned. Whether or not they all involved some form of secondary reinforcement is a moot question. It seems fair to ask in this regard, "How can one determine in any given experiment whether or not secondary reinforcers actually were present?" Merely to ask such a question suggests the *necessity for specifying—independently of the learning to be explained—some behavioral criteria which will be indicative of the operation of secondary reinforcement.* Until this is done reinforcement interpretations of latent learning or irrelevant-incentive learning are strictly unverifiable. If Spence and Kendler wish to insist upon "answers" to theoretical issues which are unequivocal, then it seems clear that until they make the specifications indicated they cannot expect disproof of their reinforcement assumption from *any* experimental design. The failure to confirm or unequivocally refute the hypothesis that learning requires reward does not arise from defects of experimental design but from untestable theoretical formulations. If it is true, as Hull has indicated, that "science has no use for unverifiable hypotheses" (**6**, p. 23) then perhaps it is time to dispense with the hypothesis that learning requires reinforcement.

I suggested in my review that we might nonetheless be interested in determining whether the hypothesis was plausible, and for this suggested especially the second of the two criteria listed above. I have explained in detail my reasons for believing that no existing reinforcement analysis satisfies this criterion. If Kendler knows of such an analysis he has not called it to our attention. It seems reasonable to conclude that reinforcement theorists are unable—despite a decade or more of concern with the problem—to formulate a consistent set of posulates from which it is possible to deduce the learning in each of these experiments. This inability itself indicates the implausibility of the hypothesis that learning requires reward.

2. Kendler's second objection was to my reasons (**20**, p. 115) for doubting that the Kendler and Mencher experiment disposed of Leeper's suggestion that discriminating perception of the irrelevant incentive is an important factor contributing to learning in type IV experiments. In the two immediately preceding paragraphs of my review I described the most directly relevant studies bearing upon the probable effects of asymmetry of intra-maze cues and forced trials. I did not argue that the avoidance of any *single* unfavorable condition would produce "positive results." Hence Grice's failure (**4**) to obtain irrelevant-incentive learning with a maze painted entirely black is not crucial to my argument. Nor is Kendler's experiment (**8**). On the other hand, the observations of Kendler and Mencher (**11**), and Meehl and MacCorquodale (**14**) and other unpublished results which I cited (**20**) clearly indicate that rats tend to develop definite preferences for black alleys and curtains over those which are white. This finding, together with that of Walker, Knotter, and DeValois (**25**), showing that animals with strong side preferences do not learn the position of the irrelevant incentive as well as those with "weak" preferences, strongly suggests that sharp asymmetry of intra-maze cues is unfavorable for such learning. Further, it seems somewhat disingenuous for Kendler to discount the possibility that a 50 per cent forcing technique may hinder learning when he (**9**) and Walker (**24**) have reported definite disturbances in some of their rats as a result of forcing to the nonpreferred side. If Kendler will read my entire review of type IV experiments as a unit, instead of considering parts out of context, he will find I have "related" many of the results he accuses me of neglecting.

Finally, Kendler's argument concerning the results of the continuity vs. noncontinuity experiments and their relevance to my criticism of the test procedures of Kendler and Mencher rests upon analogy. Grice's statement in connection with a similar experiment hinges upon the same reasoning: "If the animals had learned anything about the location of food, even if this expectation had not shown up on the [first] test trial . . . the Non-Reversal group would be expected to show a marked superiority over the Reversal group in rate of learning to select the path to food" (**4**, p. 141). My point was that any interpretation of the test series of 20 trials as providing an additional and more sensitive test of the degree of irrelevant-incentive learning has the status of a *non sequitur*. I agree that Kendler's animals gave no evidence on the first test trial of having learned the food location during the training trials. But I do not agree that comparative performances over a total of 20 test trials have any clear-cut implication regarding the existence or nonexistence of this learning. An indication of the unsuitability of the extended test treatment in this situation is provided by an unpublished experiment of mine (**21**). As in the Kendler and Mencher study, a reversal group and a nonreversal group were compared in test performance following a series of training trials involving an irrelevant incentive (in this case, only the seven initial test trials given were comparable). On the first test trial 40 per cent of the reversal group as compared with 72.7 per cent of the nonreversal group selected the food path. On the second test trial 80 per cent of the reversal group selected the correct path. Furthermore, by test trial 5 the reversal group exceeded the nonreversal group in accuracy of selecting the food path. For test trials 5 through 7 the reversal group averaged 86.7 per cent food turns, as compared with 84.8 per cent for the nonreversal group. The extended test procedure over 20 trials thus promises to obscure any learning that was evident on the first test trial.

If Kendler wishes to maintain the appropriateness of including the continuity assumption as an integral part of the Kendler and Mencher experiment, then it is obvious that he cannot consistently claim that these studies were designed only in terms of the explicit requirements of Leeper's hypothesis. It seems methodologically unsound to confound an "experimental test" of one theory by introducing procedures germane to an entirely different and opposed theory.

As to the problem of why the positive irrelevant-incentive learning experiments have often revealed only a small degree of learning, the answer I suggested (**20**, pp. 119–120) is that the constellations of experimental conditions so far sampled have been "unfavorable." I further attempted to spell out what most likely constitutes a set of "favorable conditions." Such an attempt does not in any way entail omniscience, or ability to specify precisely the necessary and sufficient conditions for such learning. Kendler apparently believes that some of the variables listed (**20**, pp. 111–120) as possibly contributing to learning in type IV experiments are unimportant. A more legitimate procedure would be for him to discuss the grounds for his belief, rather than appeal merely to Occam's Razor.

3. Although Kendler's final criticism is not relevant to my review, certain obvious confusions in his argument should be clarified. It is not true that Tolman and other cognitive theorists have failed to recognize the possibility that "the same level of performance need not imply the same level of learning." Numerous experiments cited by Tolman (**22**, pp. 39–70) on the effects of removal of reward and change in drive indicate an awareness of this possibility. Tolman has consistently stressed that the animal's demands for the available goal objects contribute heavily to his performance. It is true Tolman and others have implied that—*motivation and reward-value of the incentive being equal*—identical levels of performance should indicate identical levels of learning. Kendler is not raising a logical point at all; there is no inconsistency involved whatsoever. Kendler is really questioning (although he explicitly mentions this only in a footnote) whether the procedure of depriving animals for the same length of time necessarily results in

equal intensities of motivation. I agree that it may not, but the burden of demonstrating the inadequacy of this commonly accepted measure of drive is Kendler's.

Finally, the logic by which Kendler argues that Kanner's results (7) constitute "evidence that the *S*s in the California latent learning studies" [sic] did not have equal learning seems highly questionable. First, the generality of Kanner's results may be questioned. His experimental design was not identical with that of Tolman and Honzik —the most obvious difference is that Kanner's design provided three days of non-rewarded training as opposed to 10 days of such training in the Tolman and Honzik experiment. Schmidt (**17**), in a recently completed study, similar to Kanner's, found that when a rewarded test trial was given after varied delays his control groups (previously rewarded on all trials) were not superior in retention to his experimental groups (previously unrewarded on all but two of their training trials). A variable significantly affecting retention scores was number of previous acquisition trials (whether rewarded or unrewarded), suggesting that the foregoing discrepancy between the experiments of Kanner and Tolman and Honzik may be important. Finally, a point obscured in this whole argument is that it is not necessary to demonstrate "equal learning" in the California latent learning studies, but only to show that *some* learning occurred during the unrewarded trials.

REPLY TO MALTZMAN

Turning now to Maltzman's essay on five selected latent learning experiments, I confess that I do not know how to answer all of his criticisms. Part of my difficulty is that I cannot recognize my own statements in Maltzman's caricatures. For example, Maltzman bases much of his criticism, and presumably all of his saving score comparisons, on the argument that I claimed in my discussion of the Blodgett design that "improvement during the unrewarded period is of no importance." I deny it. The closest approximation to this that I can discover, and this removes it from the qualifying context, is: "In experiments of the Blodgett type the question of whether there is any evidence of error-reduction during the non-rewarded trials is comparatively unimportant" (**20**, p. 100). The subject of this sentence is "the question of whether there is any evidence of error reduction." In other words, I was talking about the existence or nonexistence of such evidence. To remove all possibility of misunderstanding, the meaning of my statement was precisely given in the immediately succeeding sentence: "It is not necessary that there be no evidence of a tendency toward error reduction during the unrewarded . . . trials in order for the unmanifested learning to qualify as 'latent learning.'" All this has evaded Maltzman, for he represents me as saying that *degree of improvement* during the unrewarded period is of no importance. Actually, other statements made in my review (**20**, pp. 100, 103) clearly imply that the *degree* of such improvement is relevant and important, since the very existence of latent learning in the Blodgett design is determined by statistical comparisons of error-reductions. Further, in removing my sentence from its context Maltzman distorts its predicate. I never argued, explicitly or implicitly, that the existence of improvement (let alone the degree of improvement) was unimportant in *all* contexts. I agree that the degree of improvement during "no food" trials has an important bearing on the interpretation of the results of these experiments. I had thought this was obvious from my discussion of theoretical interpretations of these experiments. The intended predicate was of course "is relatively unimportant in ascertaining the existence of latent learning." If we put the original subject with its original predicate we get something quite different from the unrecognizable (and indefensible) statement upon which Maltzman bases much of his criticism.

At times Maltzman is talking about a definition of latent learning which is unique. In discussing Blodgett's statistical comparisons using the second of the latter's criteria of latent learning, Maltzman states: "The results of this procedure are . . . of greater statistical significance than those obtained with the first criterion used. But they do not constitute a demonstration of latent learning, at least as far as Hull's theory of behavior is concerned." This statement suggests that Maltzman is using a different definition of latent learning than was discussed in my review. For him a necessary criterion of latent learning, apparently, is that Hull's theory of behavior be unable to account for it. If Hull's theory can account for it the implication is that we do not really have a demonstration of latent learning. This is a possible definition of latent learning but it is not the one I was talking about. Maltzman has confused the statistical question of whether latent learning was demonstrated with that of whether a particular theory of learning can account for the results. The former, empirical question should be clearly differentiated from the latter, theoretical question.

With these differences in our points of departure clarified, I will consider Maltzman's claims (*a*) that the studies "designated as types I and II . . . do not demonstrate latent learning," and (*b*) "that their results can be interpreted within the framework of Hull and Spence's reinforcement theory."

1. It is possible that Maltzman does not intend to question all demonstrations of latent learning in the experiments designated as

types I and II. He has ignored the results of Wallace, Blackwell, and Jenkins (26) among the former, and those of Buxton (2) and Seward (18) in the latter. Thus, at the very best, his first contention might apply to the five experiments he discusses; it could not apply, without further demonstration, to what I have called experiments of types I and II. It is questionable, however, whether such a conclusion is tenable even with reference to this truncated list of positive latent learning experiments.

The Blodgett Design (*I*). Maltzman's criticism of Blodgett's use of the one-tailed test of significance is sheer anachronism: he accuses Blodgett of violating a convention (even yet not fully established) which had not been formulated at the time of Blodgett's experiment. It seems indisputably clear that Blodgett did entertain the hypothesis of latent learning before his results were obtained. In his review of the literature the previous experiments of Lashley, Szymanski, and Simmons are described. Blodgett did not unexpectedly discover the phenomena; he was apparently checking the results of earlier experiments, and upon confirming them invented the term "latent learning." Finally, it is not unreasonable to suppose that competent research workers could use the normal probability distribution correctly even before procedural conventions were suggested for the novice. I have already alluded to Maltzman's confusion of questions of interpretation with those of existence in his discussion of Blodgett's second criterion. It is sufficient at this point to note that his criticism (that these results do not constitute a demonstration of latent learning "as far as Hull's theory is concerned") is irrelevant.

It is difficult to determine what error scores Maltzman has used in his calculation of saving scores for experiments of type I. He says such scores were obtained "by comparing the difference between the differences in performance of the control and experimental groups on the day reward is introduced and on some subsequent day." He does not tell us how much of an interval elapsed between the former and latter performances, and whether the same interval was used for both groups. If the interval consisted of several days—and apparently it did—then such a comparison would indicate the relative efficiency of learning under rewarded trials as opposed to learning under unrewarded and rewarded trials. It would not tell us anything about the degree of latent learning demonstrated. A possible measure of the latter would be a comparison of saving scores calculated directly from error scores on the day reward was introduced for each group and on the following day. Table I shows absolute and relative saving scores, calculated by this method, for the three experiments of type I discussed by Maltzman. All of these comparisons indicate superior saving for the experimental animals, and provide further demonstration of the fact that latent learning developed during the "no food" trials.

TABLE I

SAVING SCORES AS MEASURES OF LATENT LEARNING (TYPE I)

Experimental	*Group*	*Absolute Saving (Average Errors)*	*Saving in Per cent*	*Group Showing Greater Saving (Control vs. Experimental*
Blodgett	Control	.41	14	
	X-3	1.52	56	Experimental
	X-7	1.42	62	Experimental
Tolman and Honzik	Control	1.7	18	
	X-11	2.1	38	Experimental
Herb	Control	−.2*	3†	
	X-11	−2.8*	29†	Experimental

* Food in blinds.

† Denominators used in caculating relative saving scores are given by the expression 14-X, where X equals the average number of errors on the day food was introduced.

My objection to Reynolds' use of maze-wise animals (which is distinguishable from that of Blodgett) was not because of the possibility of "externalization of drive," as Maltzman says, but because of the possibility that the response or elements of the response of running the six-unit T-maze might "have been associated with reward as a result of the previous maze experience" (**20**, p. 118). A more serious criticism of Reynolds' experiment—and one which applies both to the maze-wise and the naive groups—is his use of only a single motivating drive (hunger) and a single incentive (food). Maltzman says it is difficult for him to see the relevance of this latter criticism. This is an amazing confession in view of the literature demonstrating the relationships between performance and level of drive, and between performance and quality of incentive. It seems fallacious to argue that the intensity of motivation and the quality of incentive make no difference since we are only comparing the difference in the size of error decreases in the control and experimental groups. The *reductio ad absurdum* of the proposition for which Maltzman contends is that if the animals were run under a few minutes' food deprivation and were rewarded or not rewarded with a flake of food, one should get the same significant differences in error decreases as are obtained when the motivational and incentive differentials are much greater.

The Haney Design (II). Turning to experiments of the second type, we may briefly consider the Haney (**5**) and Daub (**3**) studies. Maltzman says, "Thistlethwaite insists . . . that Reynolds' seven-day group did not meet the conditions necessary for a test of the latent learning hypothesis, because they had learned prior to the introduction of reward. The same conclusion must hold for Haney's experiment. The experimental group already learned the maze to a great extent prior to the introduction of reward." The last sentence is precisely correct. But Maltzman fails to recognize the point of Haney's experiment. The principal issue in these experiments was whether learning could occur under non-reward conditions. Maltzman forcefully concedes that it may. A significant difference, overlooked by Maltzman, between Reynolds' experiment and that of Haney is that in the former the previous maze experience was accompanied by reward.

Contrary to the impression created by Maltzman the commonly accepted test of latent learning in type II experiments is not Blodgett's set of criteria. The selection of the latter to evaluate these experiments might with some justification be termed an arbitrary selection made on the grounds that it will provide results least damaging to reinforcement theory. Maltzman does point to a difficulty, however, which is this: Even granting that the experimental animals in the Haney and Daub experiments learned without reward, why should not this knowledge of "what leads to what" remain latent on the first rewarded trial? Recent experiments by MacCorquodale and Meehl (12) suggest the answer. These investigators have confirmed the fact that rats after brief exploratory sessions in a maze may exhibit a tendency to avoid culs even though no incentive is available anywhere in the maze. Since we do not know the nature of the drive operating, these experiments do not unambiguously demonstrate that any drive-reduction (or reinforcement) was present during such trials. However, they suggest that the experimental animals of Haney and Daub may well have had "reason" to display their knowledge on the first test trial. The assumption made in the Blodgett design is that one can control motivation by manipulation of the incentives. However, whenever there are other important sources of motivation present, as seems indicated in the present experiments, one may question the control achieved by contrasting "food" and "no food" trials, and therefore the sensitiveness of Blodgett's criteria. The statistical comparisons presented by Haney and Daub, on the other hand, unequivocally demonstrate that the experimental animals, which explored the maze, showed consistent superiority over the control animals in learning the maze. One may take this as a demonstration of a latent learning which was not manifest in the pre-test performance of these animals.

2. Maltzman has greatly simplified the task of interpretation by limiting his discussion to the results of only five of the latent learning experiments. His proposed remedy in terms of habituation processes, unfortunately, does not rest upon any clear support even in the truncated results. For example, the suggested interpretation supposes that there occurs during the "no food" trials an extinction of responses incompatible with the correct responses of the training series. But

this is clearly not consistent with the results of Herb's experiment. The responses which underwent extinction in this experiment were the finally-to-be-rewarded responses; furthermore, responses incompatible with these were *increasing* in frequency during the "no food" trials. Finally, it is not clear how the postulated differential extinction (habituation) occurs. It might possibly occur through differential reinforcement. However, no source of reinforcement during the "no food" trials, or during the exploratory periods, has been identified, nor has any alleged source of reinforcement been shown to be actually present and active in the unrewarded trials of these experiments (**1**, **2**, **3**, **5**, **18**, **23**, **26**). It is therefore gratuitous to argue that these results have been shown to be interpretable within the framework of Hull and Spence's reinforcement theory of behavior.

REFERENCES

1. Blodgett, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113–134.
2. Buxton, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theor.*, 1940, **2**, No. 2.
3. Daub, C. T. The effect of doors on latent learning. *J. comp. Psychol.*, 1933, **15**, 49–58.
4. Grice, G. R. An experimental test of the expectation theory of learning. *J. comp. physiol. Psychol.*, 1948, **41**, 137–143.
5. Haney, G. W. The effect of familiarity on maze performance of albino rats. *Univ. Calif. Publ. Psychol.*, 1931, **4**, 319–333.
6. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
7. Kanner, J. H. Do non-rewarded animals learn as much as rewarded animals in the California latent learning studies? *Amer. Psychologist*, 1951, **6**, 266. (Abstract.)
8. Kendler, H. H. The influence of simultaneous hunger and thirst upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
9. Kendler, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, **40**, 265–270.
10. Kendler, H. H. Some comments on Thistlethwaite's perception of latent learning. *Psychol. Bull.*, 1952, **49**, 47–51.
11. Kendler, H. H., & Mencher, Helen C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, **38**, 82–88.
12. MacCorquodale, K., & Meehl, P. E. On the elimination of cul entries without obvious reinforcement. *J. comp. physiol. Psychol.*, 1951, **44**, 367–371.
13. Maltzman, I. The Blodgett and Haney types of latent learning experiment: Reply to Thistlethwaite. *Psychol. Bull.*, 1952, **49**, 52–60.
14. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, **41**, 372–396.
15. Meehl, P. E., & MacCorquodale, K. A failure to find the Blodgett effect, and some secondary observations on drive conditioning. *J. comp. physiol. Psychol.*, 1951, **44**, 178–183.
16. Reynolds, B. A repetition of the Blodgett experiment on "latent

learning." *J. exp. Psychol.*, 1945, **35**, 504–516.

17. Schmidt, Hans, Jr. The acquisition and retention of latent learning. Unpublished M. A. thesis, Univ. of North Carolina, 1951.
18. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.
19. Spence, K. W. Theoretical interpretations of learning. In Stone, C. P. (Ed.), *Comparative psychology* (3rd ed.). New York: Prentice-Hall, 1951.
20. Thistlethwaite, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.
21. Thisletthwaite, D. Drive discrimination and food location as factors in irrelevant-incentive learning. *Amer. Psychologist*, 1951, **6**, 287–288. (Abstract.)
22. Tolman, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
23. Tolman, E. C., & Honzik, H. C. Introduction and removal of reward, and maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 257–275.
24. Walker, E. L. Drive specificity and learning. *J. exp. Psychol.*, 1948, **38**, 39–49.
25. Walker, E. L., Knotter, Margaret C., & DeValois, R. L. Drive specificity and learning: the acquisition of a spatial response to food under conditions of water deprivation and food satiation. *J. exp. Psychol.*, 1950, **40**, 161–168.
26. Wallace, S. R., Jr., Blackwell, M. G., Jr., & Jenkins, I. Prereward and postreward performance in the "latent learning" of an elevated maze. *Psychol. Bull.*, 1941, **38**, 694. (Abstract.)

Received September 15, 1951.

CORRECTION

A. RABIN and W. GUERTIN

Michigan State College

In our paper entitled "Research with the Wechsler-Bellevue Test" in the May, 1951 issue of *Psychological Bulletin* reference was made to Dr. Kutash's research report (Reference No. 85). At the time that this study appeared it was of considerable importance in showing the necessity for caution in interchanging the Stanford-Binet and Wechsler-Bellevue tests. Through this paper as well as others the importance of the chronological age in connection with allowances for deterioration after maturity has become recognized.

At the time of the writing of this review our emphasis was on the comparability of the Stanford-Binet and Wechsler-Bellevue with respect to individuals who are in the presenescent period of life. This led us to refer to his article with such phrases as "did not control age" and "uncontrolled age factor." This presented Dr. Kutash's work in an unfavorable light and we hereby recommend that this be kept in mind in reading this review. Actually, Dr. Kutash's work showed a careful control of all relevant variables which were under consideration including that of age.

Why, after all, should we bother with all the ideas any one man has, when many of them are undoubtedly bad? Secondly, why is it that Bronfenbrenner finds a brief clinical analysis of the man so useful in understanding his theory? Thus he discovers the genesis of Rank's ideas in his relationship to Freud, of Freud's ideas in his relationship to his father, etc. This gives one the queer, and rather shocking, impression that a personality theory is a kind of gigantic projective system. And in fact this seems to be the case so long at least as we do not have data—"autochthonous" empirical data—which minimize the relative importance of the internal, personal determinants of such scientific belief systems.

Therefore, while I find Bronfenbrenner's analysis historically fascinating, the most important and interesting parts of this book for me are those written by men about their own research data. The capacity of men to invent new "symbol systems" or connotative theories has always seemed to me so infinite that new ones leave me a little cold (as does Cameron's in the present book), unless they are closely tied to actual experimental data. Nowhere is the contrast I am trying to make clearer than in Klein's chapter which reflects remarkably divided loyalties. In the first part of it, he revolts against all sorts of traditional psychological concepts, such as the nomothetic approach in general or such commonplace terms as "traits" and "habits" in particular. Thus he wants to replace a term like "perceptual trait" with "matter of course avenues" or "perceptual Anschauungen." He rejects the classical distinction of introversion-extroversion (or even extratensive-introversive) in favor of "self-inward" and "self-outward." And his discussion is so loaded with new words and distinctions that one might at this point classify him with the "connotative" theorists who seem to be interested in spinning out differences that don't make an operational difference. On the other hand his actual research discussed later in the chapter is imaginatively conceived, carefully executed, and intensely interesting. My own reaction to it is that at last someone has begun to isolate distinctive modes of perceiving the world which different individuals characteristically adopt in attempting to adjust to it. Yet somehow all the new terminology Klein introduces to explain such fascinating results seems largely unnecessary. Why give up a perfectly good commonplace term like "perceptual trait" just because someone used the term in another sense once? Thus Klein gives the impression of having one foot in the "clinical camp," where personality concepts are still rather like "projective" products, and another foot in the experimentalist camp where they are based more firmly on external, empirical facts. Perhaps only in Bruner's stimulating chapter is there evidence of a genuine synthesis of the contributions of both viewpoints into a tentative theory which avoids the "blindness" of the empiricist on the one hand and the "projections" of the clinician on the other.

Another general impression of the book is that many of the chapters represent useful condensations by the authors of longer books. Thus, Morgan summarizes some of the most important ideas in Morgan and Stellar; Beach abstracts from his recent book written in collaboration with Ford; Korzybski writes entertainingly about ideas that he has treated extensively elsewhere and manages in the process to display many of the bad language habits he so passionately objects to. Rogers, Cameron, and J. G. Miller likewise contribute condensations of longer works. In a somewhat different class are the condensations of *future* works which we will await with heightened interest from Bruner and from Frenkel-Brunswik. Finally, there are useful summaries of research conducted by others—such as Hilgard's discussion of the Ames' demonstrations and Dennis's extremely valuable summary of Piaget's research on children's ideas. Thus one can readily imagine using the book as the basis for a seminar or as the starting point for more detailed discussion of original works.

Finally, the book has a good index which, if properly used by students or in seminars, should aid in producing discussion and even integration of the differing viewpoints presented. Take the Rorschach, for example. If you look up the references under this heading in the index, you will find Dennis castigating psychologists quite properly for accepting without question the fact that a color response to an inkblot means something about emotion when the evidence supporting the association is scarcely better than that for connecting red hair with emotion, Klein going behind the Rorschach to the "perceptual attitudes" which produce responses to it, and Elsa Frenkel-Brunswik discussing the place of Rorschach in the historical development of the study of personality. The index will also show up some interesting contradictions, as, for example, under "Frequency" where there is a reference to Bruner asserting that an hypothesis will be stronger the more frequently it has been confirmed and another to Hilgard arguing that this is not always true. The student can learn much in the course of resolving such a disagreement (which in this instance is more apparent than real). In short, the book contains many new and interesting ideas presented by some of the best psychologists in the business today and should provide an excellent basis in its very variety and inconsistency for serious discussion and theorizing for a number of years to come.

DAVID C. MCCLELLAND.

Wesleyan University.

FORD, C. S., & Beach, F. A. *Patterns of sexual behavior.* New York: Harper, 1951. Pp. viii+307. $4.50.

Here is sex in cultural and evolutionary perspective. In writing this book, Ford and Beach have accomplished a difficult task. In one brief volume, they have summarized the major facts of mammalian sexual behavior, have laid the groundwork for future scientific studies and theories of sexual behavior, and have spelled out what the layman ought to know about sex. No student of behavior, comparative or physiological, social or abnormal, ought to overlook this work. No person, married or about to get married, should miss reading it.

Do not let the clear and simple writing of these authors make you believe that their book is lightweight or shallow. Although it is not peppered with references, it is a well-documented volume. It is written with the utmost scientific caution. Yet it offers interpretation. Best of all, it has a point which the scientist as well as the layman needs to learn: how to be objective about sexual behavior without depending only on cold, mass statistics.

The main scientific problem that Ford and Beach face is to determine which aspects of man's sexual behavior are the product of experience and training and which are determined by his physiological make-up. To accomplish this end, the authors evaluate sexual behavior in American society from the cross-cultural and comparative or evolutionary points of view. Practices and customs in 190 cultures are reviewed. Studies of more than thirty species of mammals from the shrew through the chimpanzee are presented. In addition are included direct experimental investigations of the roles of physiological mechanisms and the learning process in sexual behavior.

These four approaches in the analysis of sexual behavior provide excellent criteria for the evaluation and understanding of sexual practices and attitudes in American culture. In simplest terms, they amount to this: (*a*) Any practice that is widespread among the different cultures and also appears in lower mammals, particularly the subhuman primates, can be suspected of being a basic physiological pattern. (*b*) Any behavior that is practiced in some cultures and not in others and that does not appear among the lower mammals is probably the outcome of learning. When they apply these criteria, Ford and Beach follow a simple formula.

1. They take up the occurrence of the type of sexual behavior in American society and the prevalent attitude toward it. A good example is their treatment of homosexual behavior.

2. They examine other cultures which have different attitudes and practices in regard to the behavior in question. In the case of homosexuality, they report that even the most restrictive cultures have some incidence. Furthermore, some cultures are permissive and even demand that certain individuals practice homosexuality at least during some period in their lives.

3. They review the evidence on subhuman primates and subprimate mammals. Here they find that homosexual behavior does occur, that mammals are basically bisexual, and that which pattern comes out depends upon the nature of the immediate stimulation and, possibly, the past experience of the individual.

4. They examine the direct physiological evidence that the behavior is determined by the genetic makeup of the individual. In the case of homosexuality, they find that male and female hormones do not induce masculine and feminine behavior, but rather simply excite the organism to either pattern depending upon other circumstances.

5. They review studies designed to get at the role of learning. As far as homosexuality is concerned, they can show that rearing rats in isolation or segregation does not alter the incidence of homosexuality and heterosexuality compared to what is found in rats reared in cohabitation. Some evidence indicates that monkeys that are social outcasts may develop homosexuality, but it is also true that dominant males that have ready access to females may practice homosexuality in addition to heterosexual copulation.

6. Comparisons of males and females of all species and cultures are made. In this case, it is found that overt homosexual acts are much more prevalent in all of these cases in the male than they are in the female.

7. The developmental approach is also utilized. Here it is shown that homosexuality occurs much more universally in immature organisms among both animals and men than it does in the sexually mature members of the species.

8. Finally in their analysis, the authors offer an interpretation of the behavior under consideration. For homosexual behavior, they say that it is a natural type of sexual expression, dependent fundamentally on the fact that mammals, including man, have the kind of nervous systems which permit them to be aroused by masculine and feminine types of stimulation and to execute many of the components of both masculine and feminine mating patterns. It is clear that hormonal abnormalities are not typically at the basis of homosexuality, but that either hormone can serve to excite both patterns. Homosexuality occurs most commonly in sexually immature organisms, but the typical pattern for mature individuals among all animals and all cultures is heterosexual copulation. The decrease in homosexuality in the mature animal is probably largely the result of the competition offered by the heterosexual mode. Among humans, there is also the fact that many societies discourage or violently oppose homosexuality in the early training of the individual. Because of man's tremendous capacity for learning and symbolism, his homosexuality may differ from that of the lower forms. For example, completed anal copulation or the use of penis substitutes are probably impossible among lower mammals and occur only rarely in the great apes. It is also true that only among humans is there any noteworthy incidence of cases where homosexuality is more or less permanently preferred to heterosexual behavior.

The authors make the same kind of complete analysis of other aspects of sexual behavior: ways of attracting mates, the nature of foreplay and types of heterosexual stimulation, the positions and techniques of coitus, the physical and psychological circumstances under which copulation will occur, the occurrence of single and multiple sexual partnerships, masturbation, and relations with other species. Each topic is given a whole chapter, and in addition, chapters are devoted to the development of sexual behavior in the individual, the relation between sexual behavior and feminine fertility cycles, and the relative importance of hormones and central nervous mechanisms in the control of sexual behavior. An introductory chapter discusses the nature and limitations of the task undertaken in the book, and the final chapter is a good summary, viewing human sexual behavior in perspective. A good glossary is included at the end of the book.

A number of important generalizations are made in the final chapter as well as elsewhere throughout the book.

1. Changes have taken place in evolution which have released the primates, especially man, from strict hormonal control of sexual behavior.
2. The higher an animal on the evolutionary scale, the more its sexual behavior is dependent upon the cerebral cortex and the more it is influenced by learning.
3. Certain aspects of man's sexual behavior are clearly the results of his mammalian heritage: sexual play in childhood, masturbation, homosexuality, and many aspects of sexual foreplay, including mutual manipulation of the genitals with the hands and mouth and general grooming behavior.
4. Some of the similarities seen in different cultures seem to be the result of common learning experience rather than any biological mechanism common to all men. For example, while incest taboo is almost universal in human cultures, there is no good evidence for it among animals, and there are few sound biological indications against it. The authors suggest that incest regulations are designed to preserve the nuclear family group in societies by keeping it free from the dangers of sexual jealousies and conflicts.
5. It is amply clear that many of the sexual practices of man, found only in particular cultures or even only in particular individuals, are the product of social learning, very often a deliberate part of the training of the human individual.

Taken in perspective, this book is a tremendous achievement in the fields of comparative psychology and social anthropology. It is an excellent example of what these approaches can offer in other areas of human behavior as well as in the area of sexual behavior. But as the authors point out, it is only a start, really a primer. Much remains to be done, but the groundwork has been laid.

The Johns Hopkins University. ELIOT STELLAR.

MORGAN, CLIFFORD T., & STELLAR, ELIOT. *Physiological psychology* (2nd Ed.). New York: McGraw-Hill, 1950. Pp. ix+609. $5.00.

The authors of this edition have produced a distinguished successor to the original. Its coverage is broad, and for a technical book its reading style simple. In general, accuracy has not been sacrificed for readability. This edition continues to be the best available single treatment of the subject for teaching purposes. It will undoubtedly influence another generation of students exposed to this field of study.

The text has been considerably revised. Almost every area previously covered has been freshened by an evaluation of more recent experimental findings. Theoretical speculations (including those developed previously by the senior author) have been sharply reduced. A good deal of the psychophysical material has been omitted. The treatment of motor functions, color vision, and learning has been extended.

It would be easy to disagree with the authors on their choice and organization of the material. That there is room for disagreement is attributable to the breadth of the field, forcing selection on those who are courageous enough to undertake the task of presenting an objective and balanced overview of physiological psychology. The authors have been highly (but not uniformly) successful in accomplishing this task. A weakness arises out of the contradiction between their avowed main interest as the study of man (p. 502) and a heavy emphasis on the results of experiments on lower animals (in the interest of objectivity). The growth of knowledge about the main object of interest, the human organism, is thereby deferred by a preoccupation with organisms on which psychological data as intervening variables are extremely difficult to study. As a consequence, the "givens" are usually some structural insult followed by a *limited* description of affected behavior. The "in-betweens" are neglected. The temptation is strong (and frequently not resisted) to fill the void with overaged reductionist assumptions.

The reader is often led to believe that behavior modification resulting from brain extirpation means that the part extirpated is responsible for the particular behavior. If this logic were applied to the optic chiasm we should consider it responsible for vision! Obviously many other structures, and experiences as well, participate in making visually dependent behavior and experience possible. Because the authors limit the description of behavior when discussing correlated physiological events, the reader is permitted to infer that the control or responsibility for direction of an activity belongs to these events. Note, for example, the uncritical use of the words "control," "executive function," "responsible," among others, in the following quotations:

> The central nuclei of these nerves (III, IV, and VIth cranial) make up an important center for the control of eye movements. . . . (p. 24).
>
> Primary control of autonomic functions is vested in the hypothalamus (p. 43).
>
> This (the corticobulbar tract) serves the motor nuclei . . . of the cranial nerves that control movements of the eyes, face, lips, tongue and vocal cords (p. 283).
>
> . . . it is fair to conclude that area 4 is the principal executive center for motor coordination (p. 329).

Do the authors mean, for example, that the motor nuclei of the cranial nerves *direct* the movement of the eyes, face, lips, etc.? Or do they mean that these nuclei are among the structures and experiences which make all of the activities of these organs possible?

Consider also the statement: "We must ultimately understand nest building in terms of thermal changes on the nervous system" (p. 414). Apparently the authors do not mean by "ultimate understanding" a complete description of *all* of the processes involved in this activity but rather they imply a hope that some day it will be possible to demonstrate that thermal changes in the nervous system trigger nest building.

Similarly, summarizing the results of Krechevsky's work on brain mechanisms and hypotheses in rats they state: " . . . a cortical area corresponding fairly exactly with the striate (visual) area was responsible for the visual hypothesis" (p. 492). Is the production of visual hypotheses in rats a unitary psychological function emanating from the striate area or is the presence of the striate area critical for this function?

In line with the reductionist bias of the authors, the study of spinal or other reduced preparations is justified (p. 299) without cautioning the reader that the relevance of these studies to an understanding of the behavior of the intact organism is questionable. For example, the crossed extensor of the spinal cat and the reflex figure of the decerebrate cat are alleged to be basic to walking and running (pp. 301–304). The claim is made in spite of the fact that neither the conditions for eliciting them nor the actual form of the movements themselves is comparable to those for walking and running. It is highly doubtful that these reflexes have anything to do with walking and running.

Apart from disagreement with the authors on these theoretical issues, the reviewer believes this edition is superbly well done.

The City College of New York. JOSEPH E. BARMACK.

BARTLEY, S. HOWARD. *Beginning experimental psychology.* New York: McGraw-Hill, 1950. Pp. viii+483. $4.00.

This is another of the recent books aimed at an undergraduate course in experimental psychology. In particular this is a beginning book about experimental psychology for use in a lecture course without accompanying laboratory. As stated by the author, "Although laboratory work is indispensable at an early stage of your training, your first introduction to experimental psychology is well accomplished through verbal description and explanation provided by textbook and lectures."

It may be questioned whether this aim can be achieved without practical work in the laboratory. To be sure there is much debate concerning the most effective manner of teaching *experimental method* as applied to psychology. Too often the traditional course has become routinized so that students follow directions blindly with little appreciation of the nature of experimentation, the design of experiment, the application of controls, and the very essence of scientific observation. Without a laboratory course in which the student can apply these principles, however, a course in experimental psychology can become merely another subject-matter course about experimental psychology.

The author gives four characteristics of his book. Each of these will be discussed in turn.

1. "*Short enough to be covered in one quarter or one semester.*" There are 48 short chapters, each of single assignment length. In most chapters one or two typical experiments are presented in detail although some chapters follow the typical textbook form. Review questions for the student are found at the end of each chapter. Although the book is designed for a semester course, some doubling up of assignments would be necessary to fit forty-eight sessions into a typical semester with three meetings per week.

2. "*Dealing with principles by way of illustration rather than being a compendium of the findings of experimentation.*" Because of the arrangement into many short chapters the student is presented with a kaleidoscope of experimentation and methodological detail. For the beginning student the significance of these details may be lost unless he can be made to see the relevance, not only of the details of the experiment but also of the particular experiment, to psychological knowledge and theory in general. In some of these chapters there is a dearth of such background material. On the other hand, some, like Chapter 41 on "The Social Basis of Perception," are well done in this respect.

Certain of the experiments are rather complex in design and indecisive in outcome. The author does not always make clear the steps by which certain general theoretical statements are derived from experimental particulars. For example, the following is one of the conclusions of the chapter on configurational learning in the goldfish. "Perkins also saw in the behavior of the fish the exemplification of the 'law of least action.'" This is the first mention of the law anywhere in the text. There is no further statement or elucidation of the law except for a footnote as follows: "You as a student of experimental psychology, will do well to look into the concepts of least action, maximum work, etc., formulated by Wheeler. They are given in his textbook, *The Science of Psychology.*"

3. "*Representing all the major areas in psychology rather than just one or two which happen to be the central interest of the author.*" By this the author undoubtedly refers to the somewhat broader selection of fields than is usual in the traditional texts on experimental psychology. In addition to sections on History, The Nature of Experimentation, Psychophysical Methods, The Broad Features of Sensory Experience, Physiological Psychology, Comparative Psychology, Learning and Memory, and Preparedness and Adequacy, there appear five sections covering, respectively, Social Psychology, Legal Psychology, Child Psychology, Clinical Psychology, and Industrial Psychology.

4. "*Designed so that the assignments may follow the natural divisions of the material.*" It will be noted that the section headings sometimes are subject-matter headings, sometimes method headings, sometimes fields of application. Hence the material in certain chapters cuts across subject-matter headings, giving the book a "cut up" arrangement. Experiments on learning may be found in the section on Physiological Psychology, on Comparative Psychology, or on Learning and Memory itself. Studies of Perception may be found under

Social Psychology or the Broad Features of Sensory Experience. No clear organization is apparent in the succession of section headings or chapters.

The particular experiments chosen for each of the chapters are somewhat unusual. Although certain of the more typical experiments are represented, there are others that are not commonly included in textbooks. As a result, beginning students using this book are likely to be less well grounded on certain of the more classical experiments.

The author has written the book with the laudable aim of showing that experimentation is not "mere data collection" as the beginning student often thinks. He tries to alert the student to the derivation of principles from experiment by "an extended description of exemplary investigations from the several main divisions of psychology." Readers may take exception to the author's selections both of the main divisions and particular experiments. The diversity of material, lack of a clear organization and wealth of methodological detail may hamper the fulfillment of the author's aim.

CARL PFAFFMAN.

Brown University.

BUGELSKI, B. R. *A first course in experimental psychology.* New York: Holt, 1951. $3.50.

When the student begins laboratory work in psychology, he usually gets his directions from one book and studies experimental psychology from another. Bugelski's book serves both purposes, thereby achieving an integration that is often lacking. The chapter on vision illustrates the plan. Before introducing three experiments, the blind spot, depth perception, and color mixture, the author devotes 13 pages to the stimuli for vision, the size, strength and quality of visual stimuli, color names and wavelengths, and a list of variables in visual experimentation. The author cannot say much about vision in 13 pages, to be sure, but what he says helps the student do his work, gives him some terms, principles and, particularly, some information about experimental methods and their limitations. The student could get along with this book alone, though references to standard sources are mentioned.

There are 27 experiments, including three on psychophysics, five on sensory processes, three labelled motivation, five on learning, two on problem solving, and three labelled emotional behavior. Every one is worth while. According to the introduction, each experiment "has been chosen to teach at least one new technique, control, or procedural feature." The directions seem quite easy to follow; apparently they have been developed and revised in the author's classes.

Most of the experiments can be done with the facilities of small laboratories, but some require more elaborate apparatus. Ingenious construction suggestions and schematic drawings are included, and

Bugelski recommends that the students make their own apparatus. While this plan has certain advantages, it is not as inexpensive as it seems, if the instructor's time and the life of the apparatus are included in the accounting. Furthermore, the students will compare the homemade apparatus they see in psychological laboratories with the standard equipment they see in other laboratories.

The author states in his preface that he makes no attempt to conceal his biases. The sophisticated reader will agree, though the students may not catch on. Bugelski rejects meaningful learning material (p. 324), visceral signs of emotion (p. 368), personality as a whole, (p. 390), ego involvement (pp. 391–393), and so on. He prefers experiments that fit easily into the S–R framework. This book is not intended to encourage students to devise experiments in social or applied psychology. And in the last chapter the author goes out of his way to state "why some problems are neither studied nor solved." Fortunately, most students will skip this chapter, but those who read it will get a strangely limited notion of science. The truth is that scientists have always been solving problems that other scientists told them to avoid. And some students who are introduced to experimental psychology by this book may later find experimental answers for questions that the author tells them not to ask.

The style is chatty and informal. It should be easy to study. But the informality permits the author occasionally to make parenthetical comments that he probably would not make as straightforward sentences. For example, the parenthetical definition of MA on page 74 is actually wrong. And there are many passing digs at psychologists who do not follow the path of modern behaviorism. But these are minor annoyances. Looking at it from the student's point of view and allowing the author the standard number of pet peeves, this book offers an unusually convenient, reasonably accurate introduction to experimental methodology.

DONALD M. JOHNSON.

Michigan State College.

MURPHY, GARDNER. *An introduction to psychology.* New York: Harper, 1951. Pp. xvii+583. $4.25.

The primary motivation of the typical student of beginning psychology is to increase his understanding of people. Sometimes we very effectively smother this original interest, in an urgent compulsion to prove that scientific experiments have been done by psychologists. Sometimes instead we cater to his desires by concocting an innocuous course in mental hygiene or arranging a sort of group-counseling situation. But perhaps there is a more mutually satisfactory solution. Perhaps it is possible to present the facts and principles of psychology in a unified and consistent fashion, but with attention still centered upon the living, acting *individual*. This is the orientation of Murphy's *Introduction*.

It is not to be inferred that Professor Murphy has striven to reconcile the divergent pedagogical aims through a lukewarm compromise. His general psychology is quite systematic, the logical extension of the viewpoint underlying his 1947 *Personality*, and labelled fairly by Stern's term "personalistic." The very characteristics which make the position meaningful to the systematist are intended to infuse it with vitality for the student. (Unhappily, they may at the same time make the textbook unacceptable to those with differing theoretical predilections.)

It may be said, first of all, that Murphy's chief objective is to describe the actions of the whole person, rather than discuss a succession of discrete topics of traditional psychology. This is in turn reflected both by the major emphases running through the book and by the form of presentation.

1. A recurring theme is the wholeness of the organism, the ceaseless interplay of all parts and functions, the inseparability of those abstracted entities too often segregated in independent chapters. In the beginning, the direction of a cell's development depends upon its relation to other cells; much later, and more complexly, the behavior patterns commonly called "personality" are influenced by the organism's needs and by his perceptual skills, functions themselves of extended growth and spiraling sequences of learning. And so on. If this general principle is not startlingly new, at least it is more faithfully kept before the student here than in the conventional textbook.
2. A complementary emphasis is upon the notion of individuality. Evidence is constantly adduced for the uniqueness, as well as for the continuing flux, of each person's patterns of characteristics.
3. A central place is also given to the socially and culturally determined "self"—the role in which one perceives himself.

By these devices, it is hoped, the reader will be helped to concentrate upon a person's acting in a certain manner rather than upon such cold abstractions as perception or imagination.

The major goal of the book is supported admirably by the organization and style adopted. While there are 30 chapters, with many of the usual titles, there are no really isolated sections, few paragraph headings, no clear breaks in the subject matter. Rather, there is a continuous story about behavior, told almost in essay form. The writer has little preoccupation with detailed descriptions of experiments or formal enumerations of behavioral phenomena. When an observation or principle illuminates the point under consideration, it is stated, rather than being saved to fill its own niche in a fixed outline. Unless an elaboration of the principle is clearly relevant to the developing theme, it is stated quite baldly, with the reference simply noted in one of the 280-odd footnotes.

While this form and content will seem to some instructors to furnish a very satisfactory foundation in psychology, a few criticisms may be anticipated. Many of these very points, incidentally, will be interpreted by some to be strong points rather than weaknesses.

1. There is too little consideration of the scientific method in general and the experimental paradigm in particular.
2. Specific methodology and findings of psychology are inadequately covered; the writing is less tightly packed with technical material than it might be. The result may be easy reading, but the responsibility for truly representing the field remains unfulfilled.
3. Evidence is often cited too tersely, with the needed controls not at all apparent.
4. An oversimplified picture sometimes results from failure to state contradictory data and theoretical positions.
5. The many literary allusions might falsely be taken to represent the kind of observation acceptable to modern psychology, rather than serving simply as interest-getters and mnemonic devices.
6. There are a few social references so presented as to have almost the faint flavor of propaganda.

Whether these items will balance favorably or unfavorably, one attribute can be commended with confidence. In an effort to soften the technical, scientific tome by "applying psychology," certain elementary textbook writers have resorted to insipid illustrations from an imaginary collegian's life and self-conscious attempts to write in the vernacular. These all too seldom come off, and generally turn out to be only an embarrassment to both teacher and student. Murphy has the happy faculty of writing informally and at the same time unobtrusively, without being sophomoric, trifling, or overly undignified. In this respect, *An Introduction to Psychology* may serve as proof to future writers that one need not be pricked in uncomfortable succession by both horns of the style dilemma. Nor, in fact, by either horn.

FRANK W. FINGER.

University of Virginia.

GARRETT, HENRY E. *Psychology*. New York: American Book Co., 1950. Pp. ix+323. $3.00.

Consistent with his purpose, Professor Garrett has written a book for the general student "who is not planning to major in psychology, and whose acquaintance with academic psychology may begin and end with the introductory course" (p. v.). The fundamental ideas, facts, and principles of traditional psychology are presented and interpreted in terms of the familiar experiences of college students. We recognize the difficult task of such a presentation, avoiding the easy drift toward the static summaries of psychological experiments and theories or to the "social and moral preachments no matter how vital these issues may be to the instructor personally" (p. vi). Perhaps the over-all text would be more attractive and interesting reading if the author had not remained so religiously close to the classical topics and evidence of general psychology. The lonely one-page treatment of "Tasting and Smelling" contributes little to the main theme of the chapter on "Perception and Observing," which in turn almost ignores the highly stimulating (at least to the student) recent research on perceptual dynamics and phenomenology.

The nine chapter headings illustrate the traditional coverage: What Psychology is About; Heredity and Growth; Motivation and Adjustment; Perception and Observation; Learning: Principles and Methods; Learning and Studying Effectively; Intelligence; Individual Differences and Aptitudes; Personality. Following each chapter are the usual "Questions and topics for discussion." The author has done a superior job in presenting questions which require the student to draw the implications and applications of the principles developed in the chapter.

It is difficult to be briefly systematic and the author chose to curtail discussion of several topics (emotions, sensory processes, thinking, and social factors in personality, etc.). A two-page treatment of emotions, for example, is presented under the heading of "Psychological Motives," and except for the heading, practically no attempt is made to show the systematic importance of emotions in human experience and behavior. Chapters dealing with personality deviations, interpersonal relations, and statistics are purposefully omitted since they are believed to belong more appropriately to specific follow-up courses. Learning, however, is extensively treated in two complete chapters.

The one-course psychology student for whom this book was written will have the opportunity to learn the basic contributions of scientific general psychology. This material is presented in an honest and, we believe, successful attempt to make it meaningful *and* useful. The 323 pages in the text represent the short end of the psychology textbook spectrum. Some of the color we have seen in recent beginning texts has been omitted and it is quite possible that if read in the evening the material would sometimes appear rather gray. We are, however, pleased to note that this well-known author did not sacrifice a "solid" psychology for popular demand.

S. C. Ericksen.

Vanderbilt University.

Pronko, N. H., & Bowles, J. W., Jr. *Empirical foundations of psychology.* New York: Rinehart, 1951. Pp. xvi+464. $3.75.

"The primary motivation toward assembling the works included here," say the authors, "was a sensitivity to writings that they believe presage a revolution in the field of psychology comparable to the Einsteinian one in physics." Presumably, then, the selections included—

some originals of their own, some rewritten from the works of others—may be a harbinger of an "age of psychology." The chief aim, they assert further, is to "provoke the student to think about fundamental problems in psychology which have often been treated complacently or as closed matters." The selections are intended to serve as "whetstones against which the student may sharpen his own intellectual tools."

Swept enthusiastically along by spiralling promises, his appetite further sharpened by assertions that the seasoning of the present psychological dish is attained largely by empirical or operational material (including both man's and nature's experiments), the reader hastens from the hors d'oeuvre of the preface to the pièce de résistance which the authors claim to have assembled.

And what does he find to appease his stimulated hunger? The content of the book, organized under the major topic headings given below, fits the form of the orthodox textbook and in that sense facilitates use of the book as parallel supplementary reading. The sub-topics used to develop the main topics represent careful cullings of unusual material from scattered sources and often provide uncommon insights into common phenomena less well treated in conventional texts. In many instances these sub-topics involve accounts of experiments which in standard texts receive only brief mention (or none at all). For the beginning student it may be that the key to intellectual stimulation comes from such detailed presentation of experimental work.

As an example of selections, the section on Social Behavior contains such items as (*a*) The social psychological response, (*b*) Culture, (*c*) Food: The story of coffee, (*d*) Dress and fashion: White man's fashions, (*e*) Superstitions of school children, (*f*) Voodoo death, (*g*) Sexual behavior and social level, (*h*) Cultural factors in gesture, (*i*) Folie à deux, (*j*) Contingential responses, (*k*) Idiosyncratic behavior.

Though reasonably well satisfied with the material presented, the reviewer nevertheless wishes for some rationale to account for the choices made, the omissions, the varying emphasis, and particularly for some insight into the sequence in which the topics are arranged. After the first five sections very little logic appears to govern the order of the remainder. For instance, the phenomena of learning are covered in section 15, the next to the last, while compelling arguments would favor its appearance much earlier as a basis and explanation for much of the material presented before this section but dependent upon it for clarification. The authors may have anticipated this critical comment when they remark in the preface that no attempt had been made to assemble a systematic text and, further, that the aim was "to bring together scattered nuggets from psychological literature." Nuggets, like pearls on a string, may increase in value if properly related to each other. And the psychologist, like any other scientist, is interested in *related* facts.

Reaction to the book is determined perhaps less by the material per se than by its failure to satisfy the expectations aroused. The heady assertion that the works assembled here "presage a revolution in the field of psychology comparable to the Einsteinian one in physics" can be accepted only if one (*a*) underestimates Einstein's contributions, and (*b*) overestimates the efforts of Pronko and Bowles. Few psychologists will wish to do so. The book is a good illustration of the disappointment that can result when an otherwise acceptable product is oversold. The work has merit and deserves an audience—but not a place of honor in the Einsteinian hall of fame.

GEORGE F. J. LEHNER.

University of California, Los Angeles.

ARNOLD, WILHELM. *Das Raumerlebnis in Naturwissenschaft und Erkenntnistheorie.* (The experience of space in natural science and epistemology.) Nürnberg: Sebaldus-Verlag, 1949. Pp. 176.

About one-fourth of this monograph is devoted to a historical-theoretical survey of the problem of space perception. This is followed by the description of a new device for testing and measuring acuity of stereoscopic vision. The apparatus is in a way an extension and modification of the one used by Helmholtz in the experiment where three vertical needles are placed in a plane perpendicular to the line of vision. By moving the middle needle out of the plane and towards or away from the observer one is able to determine the threshold values of depth perception. The author's apparatus uses a stereoscope with two sets of three threads in a plane perpendicular to the line of vision. In this device the threads remain in the same plane and perception of depth is achieved through binocular parallax by moving one of the middle threads laterally. It is now also possible to permit the observer to move the middle thread in the other set until all three threads are again perceived as lying in the same plane. The chief advantage of this device lies in the absence of additional cues of depth perception which might be added if a thread is moved closer towards the observer or away from him. Proper measuring devices permit quantitative observations.

The rest of the monograph contains a report of readings obtained at the Psychological Institute of the University of Munich and a discussion of the practical value of the new apparatus. It is suggested that among other things it might be used for training in depth perception and for the determination of individual differences in acuity of depth perception which play a role in certain professions and in piloting a plane.

There are 80 general references to the experience of space taken from the fields of philosophy, psychology, esthetics and physics; and 408 specific references to binocular depth perception.

KARL F. MUENZINGER.

University of Colorado.

Cattell, Raymond B. *Personality: A systematic, theoretical and factual study*. New York: McGraw-Hill, 1950. Pp. xii+689. $5.50.

The aim of this book in the author's words is "to treat personality study by the same scientific standards as are maintained in experimental psychology and to integrate it with research and systematic psychological concepts." This is a very ambitious aim since as the author points out the hypotheses which at present "have a close relationship to facts are restricted in range." It is offered as an antidote to the "rank and wanton pedantry of 'schools' more interested in the social pomp of an all-embracing theory than in a sober desire for truth," a chin-leading statement in a book which itself has systematic theoretical aims.

Starting with the definition of personality as "that which permits a prediction of what a person will do in a given situation" and the truism that scientific advance depends upon exact measurement and exact measurement depends upon exact description, the early chapters deal with varieties of description and measurement. They include a discussion of the devices for obtaining personality data—life history and behavior ratings, self-inventories and tests; a discussion of various techniques of covariation analyses of relationships of persons on tests, tests on persons, and tests on occasions. They include a discussion of surface traits and "source" traits in their various subdivisions or modalities and the statistical determination thereof. In short, the early chapters are a condensation of the author's *Description and Measurement of Personality* (1946). In the reviewer's opinion the condensation often results in a loss of clarity and gives the impression of arbitrariness.

In the remaining chapters an attempt is made to pull together the factorial findings of research literature where validation had been established by three or more independent studies and to point out where further research is needed. Three chapters deal with biological influences in personality under the headings of Inherited, Constitutional Influences, and Psychosomatics. Four chapters are devoted to psycho-dynamics (I. The Intrinsic Structure; II. The Structure of Innate Drives; III. The Structure in the Adjustment Process; IV. The Self—Its Integration, Adaptation and Adjustment). Four chapters are concerned with personality and the social matrix (I and II are on the family: Structure and Setting is the first and Its Relation to Child Personality is the second; III. Group Dynamics; IV. Specific Social Organs). Three chapters concern the abnormal and unadapted personality (I. The School Misfit, the Delinquent, the Criminal; II. The Neurotic; III. The Psychotic). Two chapters are on the life stages in personality (I. Conception to Puberty; II. Adolescence, Maturity and Old Age). A final chapter is devoted to Principles of Personality Formation in which 17 principles of personality are enunciated.

From this sweeping coverage of methods of data collection, of factorial analysis (R, Q, and P), of lists of surface traits derived from clusters of correlations and source traits derived from factoring, constitutional and environmental mold traits in the modalities of ability traits, temperament traits and dynamic traits (goal and sub-goal seeking), and from the final list of principles of personality formation, this reviewer emerged with no apperceptive coherence or theoretical clarity. The reviewer shares the author's belief that to build substantially in this complex field one must start with empirical descriptive facts, discover relationships, develop constructs which will lead to more precise measurements and higher level organizing constructs, better designed research, and eventually to established principles. Perhaps the data in this field are still too fragmentary, too crude, too unplanfully gathered, and too embedded in semantic imprecision and confusion to permit of systematic coherence, but the reviewer felt that the book did not approximate a return to its original definition of personality as "that which permits a prediction of what a person will do in a given situation." The book, however, should be read by anyone attempting to do serious research in the field of personality, since it contains shrewd and critical evaluations of productive methods of analysis, much important research material, and many suggestions for fruitful research attack.

University of California. JEAN WALKER MACFARLANE.

CATTELL, RAYMOND B. *An introduction to personality study.* New York: Longmans, Green, 1950. $1.50 (text); $2.00.

This abridged version of Cattell's recently published text, *Personality: A Systematic, Theoretical and Factual Study* must be viewed in two different lights: first as an introduction to personality study for the sophisticated layman and second as a presentation of the author's conceptualization of personality. Despite Cattell's belief that he has in the main presented a consensus of psychological opinion, he generally states only his own position, and unfortunately he frequently treats alternative positions with open contempt. Phrases such as "we cannot . . . pause to refute every extravagant view that has ever been put forward on human motive" and labels like "extremist" or "more enthusiastic than discreet" are too frequent. As a presentation of Cattell's position, however, the book is clear and readable, logically organized, and well worth reading.

This position can be summarized only briefly. Cattell accepts instincts defined in much the same terms as McDougall used. Several methods for the empirical discovery of instincts are described, but the author clearly favors factor analytic procedures. A basic erg is subsidiated by various symptomatic attitudes which describe the means by which the need is satisfied. The different attitudes which subsidiate the same erg are expected to have a high correlation which will enable them to be identified by factor analysis. At the same time the basic doctrine that different subsidiations of the same erg may substitute for each other and be alternative outlets, is accepted. The existence of ten or twenty basic ergs is reconciled with the psychoanalytic position by means of a postulated maturation process. The psychoanalytic instincts, through differentiation and fusion, mature into the basic ergs of the adult.

The process of adjustment is conceptualized in terms of six adjustment crossroads: (1) stimulated drive, (2) initial frustration, (3) failure of aggression, (4) anxiety and renunciations, (5) the fate of repression, and (6) neurotic reactions. At each crossroad there are various possible sequels, one of which leads to the next crossroad while the others lead back to previous crossroads or to some sort of successful adjustment. This conceptualization is an attempt to clarify the various psychoanalytic mechanisms and the viscissitudes of the instinct. It may not be entirely adequate as a theory, but it is an excellent presentation of a complex set of relationships which is frequently difficult for students to understand.

In terms of this theoretical position, Professor Cattell discusses in a chapter each, abnormal personality, psychosomatics, personality and the culture pattern, and the life course of personality. In two chapters he presents a very readable account of factor analysis and the results of his factor analytic description of personality.

The conceptualization which Cattell builds seems to the reviewer to be rather more a diagrammatic presentation of the points of view which the author has adopted than an integration of these points of view into a single coherent theory. The reconciliation of McDougall, psychoanalysis, and factor analysis is more verbal than real, because the logical difficulties are not faced squarely. For example, the doctrine that different attitudes which subsidiate the same erg are substitutable alternatives for each other implies that in a group of people with the erg at constant strength, the correlation between attitudes subsidiating the same erg would be negative. It is only when the strength of the erg varies that there would be a factor tending toward positive correlation between attitudes subsidiating the same erg. These two opposing tendencies would apparently make factor analysis an insensitive tool for revealing the existence of a common ergic root to a group of attitudes.

Cattell states that the psychoanalytic instincts differentiate and fuse in the course of development to become the basic ergs of the adult. Because the meaning of differentiation and fusion is not clear, the distinction between this process and subsidiation is unclear. Presumably the psychoanalytic theory would demand that the relation between Cattell's ergs and the instinctual drives be one of subsidiation. Thus the crucial issue is left unclear.

In summary, the book makes interesting reading. It is an attempt to do something which badly needs doing, to integrate the various points of view about personality. In view of the heroic proportions of the task it is not surprising that different readers will feel that one or another aspect of the task has been neglected. Nevertheless, to this reader it has achieved only a superficial unity.

ALFRED L. BALDWIN.

University of Kansas.

YACORZYNSKI, G. K. *Medical psychology. A basis for psychiatry and clinical psychology.* New York: Ronald Press, 1951. Pp. 535. $6.00.

In the author's words "The purpose of this book is to present an integrated approach to the understanding of human behavior as a foundation for the study of psychiatry and clinical psychology. Students of medicine require a fundamental background of this kind for their later study of diagnostic methods, psychosomatic medicine, neuroses and psychoses, and finally therapy. Similarly, students majoring in clinical psychology stand in need of an integrated approach oriented primarily toward an understanding of abnormal behavior, without the bias of any particular school."

The book is divided into three parts: "Basic Psychobiological Principles," "Inheritance and Maturation," and "Integration: Structure and Structuralization of Personality." Neither giving the headings of these parts nor the titles of individual chapters conveys either the wealth of material contained in this book or the earnest, and for the most part, successful attempt to integrate theory and fact, psychology and biology. The scope of this book is such that this reviewer has concluded that it could be used as a text in an abnormal psychology course, in an intensive introductory psychology course, and as a good reference for clinical courses. Perhaps its most meritorious aspect is the refreshing, clear and non-superficial manner in which the psychological and biological aspects of functioning are related. The psychology student's perspective upon the problems in his field will be enlarged by the author's biological orientation (and background); similarly, the medical student will get a much-needed orientation toward what is meant by the "total" individual.

SEYMOUR B. SARASON.

Yale University.

BARKER, ROGER G., & WRIGHT, HERBERT F. *One boy's day: A specimen record of behavior.* New York: Harper, 1951. Pp. ix+435. $3.50.

Because of the nature of this publication, it is not possible to review *One Boy's Day* critically. The book contains a record of 14 hours of consecutive observation of a seven-year-old, midwestern, small-town boy, at home, at school, and at play. Each of eight observers recorded in one-minute intervals the objective behavior of Raymond Birch in his natural habitat and interpretive comments regarding that behavior. Thus we see that on April 26, 1949 at 9:42 A.M., Raymond acted in the following manner in his second-grade classroom:

9:42 Raymond hurried out of the music room and back to his classroom, not boisterously, but as though he were eager to return.

"He walked briskly to his desk, on which his May basket materials lay just as he had left them.

> "At once he put his whole heart into working on his May basket and getting the paper fastener through the hole at the end of each paper strip.

The theoretical framework within which the observations were collected is not included in this volume. *One Boy's Day* contains, in effect, only raw data. Obviously we cannot evaluate raw data without knowing the frame within which the data were gathered. The reader, therefore, is left with such questions as why these kinds of data and not others, how can the data be used, etc., etc. It is unfortunate, in this reviewer's opinion, that the authors did not enlarge the volume to include their formulation of the meaning and task of psychological ecology.

CELIA BURNS STENDLER.

University of Illinois.

BOOKS AND MATERIALS RECEIVED

Balz, Albert G. A. *Cartesian studies.* New York: Columbia Univ. Press, 1951. Pp. vi+328. $4.50.

Bekker, Sarah M., & Harris, Catherine E. *Research relating to children: An inventory of studies in progress, reported August–December 1950 to the Clearinghouse for Research in Child Life.* Washington, D. C.: Federal Security Agency, Social Security Administration, Children's Bureau, 1950. (Suppl. No. 3.) Pp. 128.

Brav, Stanley R. (Ed.) *Marriage and the Jewish tradition: Toward a modern philosophy of family living.* New York: Philosophical Library, 1951. Pp. xiii+218. $3.75.

Bennett, Wendell C. *Area studies in American universities.* New York: Social Science Research Council, 1951. Pp. x+82.

Buros, Oscar Krisen. *Statistical methodology reviews 1941–1950.* New York: John Wiley, 1951. Pp. 457. $7.00.

Cameron, Norman, & Magaret, Ann. *Behavior pathology.* Boston: Houghton Mifflin, 1951. Pp. xvi+645. $5.00.

Committee on Education, Training and Research in Race Relations of the University of Chicago. *Inventory of research in racial and cultural relations*, 1951, **3**, Bull. No. 4.

Cory, D. W. *The homosexual in America: A subjective approach.* New York: Greenberg, 1951. Pp. xvii+326. $4.00.

Dewey, Richard, & Humber, W. J. *The development of human behavior.* New York: Macmillan, 1951. Pp. xiv+672. $5.50.

Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler R. W. *Intelligence and cultural differences: A study of cultural learning and problem-solving.* Chicago: Univ. of Chicago Press, 1951. Pp. xii+388. $5.00.

Harmon, Francis L. *Principles of psychology.* (Rev. ed.) Milwaukee, Wis.: Bruce, 1951. Pp. xi+656. $4.25.

Harrower, M. R., & Steiner, M. E. *Large scale Rorschach techniques: A manual for the group Rorschach and multiple choice tests.* (2nd ed.) Springfield, Ill.: Charles C Thomas, 1951. Pp. xx+353. $8.50.

Hayes, Cathy. *The ape in our house.* New York: Harper, 1951. Pp. vi+247. $3.50.

Hofstaetter, Peter R. *Die Psychologie und das Leben.* Wein: Humboldt-Verlag, 1951. Pp. 287.

Jahoda, Marie, Deutsch, Morton, & Cook, Stuart W. *Research methods in social relations: With especial reference to prejudice. Part One: Basic processes. Part Two: Selected techniques.* New York: Dryden

Press, 1951. Vol. I: Pp. x+421, $3.75. Vol. II: Pp. x+423–759, $3.75. ($6.00, set.)

JANIS, IRVING L. *Air war and emotional stress: Psychological studies of bombing and civilian defense.* New York: McGraw-Hill, 1951. Pp. ix+280. $5.00.

LUDLOW, WILLIAM L. *A syllabus and a bibliography of marriage and the family.* New Concord: Radcliffe Press, 1951. Pp. 309. $4.00.

MARX, MELVIN H. *Psychological theory: Contemporary readings.* New York: Macmillan, 1951. Pp. xi+585. $5.00.

MEAD, MARGARET. *Soviet attitudes toward authority.* New York: McGraw-Hill, 1951. Pp. 148. $4.00.

MELZER, JOHN HENRY. *Functional logic.* Dubuque, Ia.: Wm. C. Brown, 1951. Pp. vii+110. $3.75.

MILLER, GEORGE A. *Language and communication.* New York: McGraw-Hill. Pp. xiii+298. $5.00.

OSBORNE, ERNEST G. *The family scrapbook.* New York: Association Press, 1951. Pp. xv+457. $3.95.

PIÉRON, HENRI, FESSARD, ALFRED, & FRAISSE, PAUL. *L'année psychologique: Cinquantième année, volume jubilaire hommage à Henri Piéron.* Paris: Presses Universitaires de France, 1951. Pp. xvi+718. 2.400 frs.

QUEENER, E. L. *Introduction to social psychology.* New York: William Sloane, 1951. Pp. xiv+493. $4.25.

SANFORD, CHARLES W., HAND, HAROLD C., & SPALDING, WILLARD B. (Eds.). *The schools and national security: Recommendations for elementary and secondary schools.* (Illinois Secondary School Curriculum Program. Bull. No. 16.) New York: McGraw-Hill, 1951. (Springfield, Ill.: Office of Public Instruction, 1951.) Pp. xiv+292.

SCHUSTER, GEORGE. *Christianity and human relations in industry.* London: Epworth Press, 1951. Pp. 128. 6*s*, 6*d*.

SCHWEBEL, MILTON. *The interests of pharmacists.* New York: King's Crown Press, 1951. Pp. xii+84. $1.75.

SEARS, PAULINE SNEDDEN. *Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. Psychological Monographs* No. 323 (Vol. 65, No. 6). Washington, D. C.: American Psychological Assn., 1951. Pp. iv+42. $1.50.

STONE, CALVIN P. (Ed.). *Comparative psychology.* (3rd ed.) New York: Prentice-Hall, 1951. Pp. xvii+525. $6.00.

THOMPSON, LAURA. *Personality and government: Findings and recommendations of the Indian Administration Research.* Mexico, D. F.: Ediciones del Instituto Indigenista Interamericano, 1951. (Institute of Ethnic Affairs, Inc., 810 18th St., N. W., Washington 6, D. C.) Pp. xviii+229. $2.00.

VALENTINE, C. W. *Psychology: And its bearing on education.* New York: Philosophical Library, 1951. Pp. xvi+674. $6.00.

ZILBOORG, GREGORY. *Sigmund Freud: His exploration of the mind of man.* New York: Charles Scribner's, 1951. Pp. 132. $2.00.

Vol. 49, No. 3 May, 1952

Psychological Bulletin

THE NATURE AND MEASUREMENT OF MEANING[1]

CHARLES E. OSGOOD
University of Illinois[2]

No. 2 March 1952

The language process within an individual may be viewed as a more or less continuous interaction between two parallel systems of behavioral organization: sequences of central events ("ideas") and sequences of instrumental skills, vocalic, gestural, or orthographic, which constitute the communicative product. A communicator vocalizes, "It looks like rain today; I'd better not wash the car." This output is a sequence of skilled movements, complicated to be sure, but not different in kind from tying one's shoes. Even the smallest units of the product, phonetic elements like the initial "l"-sound of "looks," result from precisely patterned muscle movements. The organization of these movements into word-units represents skill sequences of relatively high predictability; certain longer period sequences involving syntactical order are also relatively predictable for a given language system. But execution of such sequences brings the communicator repeatedly to what may be called "choice-points"—points where the next skill sequence is not highly predictable from the objective communicative product itself. The dependence of "I'd better not wash the car" upon "looks like rain today," the *content* of the message, reflects determinants within the semantic system which effectively "load" the transitional probabilities at these choice-points.

It is the communicative product, the spoken or written words which follow one another in varying orders, that we typically observe. Since we are unable to specify the stimuli which evoke these communicating reactions—since it is "emitted" rather than "elicited" behavior in Skinner's terminology (**97**)—measurements in terms of rates of occurrence and transitional probabilities (dependence of one event in the stream upon others) are particularly appropriate (cf., Miller, **76**). Interest may be restricted to the lawfulness of sequences in the observable communicative product itself, without regard to the semantic parallel. This is traditionally the field of the linguist, but even here it has proved necessary to make some assumptions about meaning (cf., Bloomfield, **4**). On the other hand, one may be specifically interested in the semantic or ideational level. Since he is presently unable to observe this level of behavior directly, he must use observable characteristics of the communicative product as a basis for making inferences about what is going on at the semantic level. He may use sequential orderliness in the product to draw conclusions as to semantic orderliness in the speaker's or writer's mediation processes (i.e., which "ideas" tend to go together in his thinking with greater than chance probabilities). Or he may wish to study the ways in which central, semantic processes vary from concept to concept, from person to person, and so on. It is the problem of measuring meaning in this latter sense which will be discussed in the present paper.

---

[1] The research on which in part this report is based is being supported jointly by the University Research Board of the University of Illinois and the Social Science Research Council. Grateful acknowledgment is made to both institutions for their assistance.

[2] Institute of Communications Research and Department of Psychology.

Before inquiring into the measurement of the meaning of signs, for which there are no accepted, standardized techniques available, we may briefly mention certain fairly standard methods for measuring the comparative strength of verbal habits. Thorndike and his associates (**102, 103**) have made extensive *frequency-of-usage counts* of words in English; that this method gets at the comparative habit strengths of word skill sequences is shown by the fact that other measures of response strength, such as latency and probability within the individual (Thumb and Marbe, **106**; Cason and Cason, **19**), are correlated with frequency-of-usage. Zipf (**117, 118**, and elsewhere) has described innumerable instances of the lawfulness of such habit-strength measures. Whether samples be taken from Plautine Latin, newspaper English, or the English of James Joyce in his *Ulysses*, a fundamental regularity is found, such that frequency of occurrence of particular words bears a linear relation to their rank order in frequency, when plotted on double-log paper (Zipf's Law). Measurement of flexibility or diversity in communicative products is given by the *type-token ratio* (TTR): with each instance of any word counting as a token and each different word as a type, the greater the ratio of types to tokens the more varied is the content of a message. This measure can be applied comparatively to different forms of material, different kinds of individuals, and so forth (cf., Carroll, **16, 17**; Johnson, **45**; Chotlos, **20**), provided the sizes of samples are constant. One may also count the ratios of adjectives to verbs (Boder, **5**), the frequencies of different pronouns, intensives, and so forth (cf., Johnson, **45**).

Although the above measures get at the comparative strengths of verbal skill sequences per se (i.e., without regard to meaning), this is not a necessary restriction. Frequency counts of this type can be applied to *semantic habit strengths* as well. Skinner (**96**) has shown that a similar lawfulness applies to the frequencies of "free" associations in the Kent-Rosanoff tests. When frequencies of particular associates to given stimulus words for a group of subjects are plotted against their rank order in frequency, a straight-line function on double-log paper results (Zipf's Law). In other words, associations at the semantic level appear to be organized in such a way that few have very high probability of occurrence and many have low probabilities of occurrence. Bousfield and his collaborators (**7, 8, 9, 10, 11**) have described a *sequential association method* for getting at comparative semantic habit strengths. When subjects associate successively from the same "pool," e. g.,"names of four-legged animals," (*a*) the rate of successive associates shows a negatively accelerated curve, (*b*) varying in its constants with certain characteristics of materials and subjects, (*c*) the order of appearance of particular associates in individuals being predictable from the frequency of usage in the group, and (*d*) distortions in the function being related to particular transitional probabilities among associates, i.e., clustering. Useful though these measures are for many purposes, they do not get at meaning. The fact that "dog" has a higher probability of occurrence in sequential association than "otter" says nothing whatsoever about the differences in meaning of these two signs.

An extensive survey of the literature fails to uncover any generally accepted, standardized method for measuring meaning. Perhaps it is because of the philosophical haziness of this concept, perhaps because of the general belief that "meanings" are infinitely and uniquely variable, or perhaps because the word "meaning" as a construct in our language connotes mental stuff, more akin to "thought" and "soul" than to anything observable—for some combination of reasons there has been little attempt to devise methods here. Nevertheless, whether looked at from the viewpoints of philosophy or linguistics, from economic or sociological theory, or—interestingly enough—from within the core of psychological theories of individual behavior, the nature of meaning and change in meaning are found to be central issues. The proposals to be made in the latter portion of this paper are part of a program aimed at the development of objective methods of measuring meaning. Beyond obvious social implications, it is felt that this direction of research is a logical extension of scientific inquiry into an area generally considered immune to its attack.

THEORIES OF MEANING

Not all stimuli are signs. The shock which galvanizes a rat into vigorous escape movements usually does not stand for anything other than itself, nor does the pellet of food found at the end of a maze, nor a hammer in one's hand or a shoe on one's foot. The problem for any meaning theorist is to differentiate the conditions under which a pattern of stimulation is a sign of something else from those conditions where it is not. This certainly seems simple enough, yet it has troubled philosophers for centuries. By stating the problem somewhat formally, the chief differences between several conceptions of the sign-process can be made evident: let

$\dot{S}$ = object = any pattern of stimulation which evokes reactions on the part of an organism, and

$\boxed{S}$ = sign = any pattern of stimulation which is not this $\dot{S}$ but yet evokes reactions relevant to $\dot{S}$—conditions under which this holds being the problem for theory.

The definition of $\dot{S}$ is broad enough to include any pattern of stimulation which elicits any reaction from an organism. Although one usually thinks of "objects" as those things denoted by signs, actually any pattern of stimulation—a gust of northerly wind against the face, the sensations we call "belly-ache," the sensations of being rained upon—is an "object" at this level of discourse. One sign may be the "object" represented by another sign, as when the picture of an apple is called "DAX" in certain experiments. The definition of $\boxed{S}$ is purposely left incomplete at this point, since it depends upon one's conception of the nature of the sign-process.

We may start a logical analysis of the problem with a self-evident fact: *the pattern of stimulation which is the sign is never identical with the pattern of stimulation which is the object.* The word "hammer" is not the same stimulus as is the object hammer. The former is a pattern of sound waves having characteristic oscillations in frequency and intensity; the latter, depending upon its mode of contact, may be a visual form having characteristic color and shape, a pattern of tactual and proprioceptive sensations, and so on. Similarly, the buzzer in a typical rat experiment is not identical as a form of stimulation with the shock which it comes to signify. Yet these signs—the word "hammer" and the buzzer—do elicit behaviors which are in some manner relevant to the objects they signify, a characteristic *not* shared with an infinite number of other stimulus patterns that are *not* signs of these objects. In simplest terms, therefore, the question is: *under what conditions does something* *which is not an object become a sign of that object?* According to the way in which this question is answered we may distinguish several theories of meaning.

Mentalistic View

The classic interpretation derives directly from the natural philosophy of Western culture, in which the dualistic connotations of language dictate a correlation between two classes of events, material and nonmaterial. Since meanings are obviously "mental" events and the stimuli representing objects and signs are obviously "physical" events, any satisfying theory of meaning must specify interrelation between these levels of discourse. At the core of all mentalistic views, therefore, we find an "idea" as the essence of meaning; it is this mental event which links or relates the two different physical events, sign and object. The word "hammer" gives rise to the idea of that object in the mind; conversely, perception of the object hammer gives rise to the same idea, which can then be "expressed" in appropriate signs. In other words, *something which is not the object becomes a sign of that object when it gives rise to the idea associated with that object.* Probably the most sophisticated expression of this view is given by Ogden and Richards (82) in their book, *The Meaning of Meaning.* Most readers will recall their triangular diagram of the sign-process: the relation between symbol and referent (the base of their triangle) is not direct but inferred, mediated through mental "thought" or "interpretation" (the third corner of their triangle).

Substitution View

Naive application of Pavlovian conditioning principles by early behaviorists like Watson led to the theory that signs achieve their meanings simply by being conditioned to the same reactions originally made to objects. This, in essence, is the view one encounters in many introductory texts in general psychology. An object evokes certain behavior in an organism; if another pattern of stimulation is consistently paired with the original object, it becomes conditioned to the same responses and thus gets its meaning. The object is the unconditioned stimulus and the sign is the conditioned stimulus, the latter merely being substituted for the former. The definition of the sign-process here is that *whenever something which is not the object evokes in an organism the same reactions evoked by the object, it is a sign of that object.* The very simplicity of this theory highlights its inadequacy. Signs almost never evoke the *same* overt responses as do the objects they represent. The word FIRE has meaning to the reader without sending him into headlong flight. Nevertheless, this represents a first step toward a behavioral interpretation of the sign-process.

Meaning as "Set" or "Disposition"

In a monograph entitled *Foundation of the Theory of Signs* (77), Charles Morris, a semiotician working in the tradition established by Peirce and other American pragmatists, proposed a formula for the sign-process which avoids the pitfalls of substitution theory but seems to step backward toward the mentalistic view. In essence he states that signs achieve their meanings by eliciting reactions which "take account of" the objects signified. The sign "hammer" may evoke quite different responses from those evoked by the object signified, but these responses must have the character of being relevant to the object. The response made to the sign is called the "interpretant" which mediately takes account of the object signified. But it would seem that this process of "taking account of" is precisely what needs elucidation.

During the period intervening between this monograph and his recent book, *Signs, Language and Behavior* (78), Morris studied with two prominent behavior theorists, Tolman and Hull. The effects of this immersion in learning theory are evident in his book, which is a pioneer attempt to reduce semiotic to an objective behavioral basis. He states that "if anything, *A*, is a preparatory stimulus which in the absence of stimulus-objects initiating response-sequences of a certain behavior-family causes a disposition in some organism to respond under certain conditions by response-sequences of this behavior-family, then *A* is a sign" (p. 10). Reduced to its essentials and translated into our terms, this becomes: *any pattern of stimulation which is not the object becomes a sign of that object if it produces in an organism a "disposition" to make any of the responses previously elicited by that object.* There is no requirement that the *overt* reactions originally elicited by the object also be made to the sign; the sign merely creates a disposition or set to make such reactions, actual occurrence depending upon the concurrence of supporting conditions.

Beyond the danger that "dispositions" may serve as mere surrogates for "ideas" in this theory, there are certain other difficulties with the view as stated. For one thing, Morris seems to have revived the substitution notion. The sign is said to dispose the organism to make overt response-sequences of the *same* behavior-family originally elicited by the object. But is this necessarily the case? Is my response to the word "apple" (e.g., free-associating the word "peach") any part of the behavior-family elicited by the object apple? For another thing, Morris' formulation fails to differentiate sign-behavior from many instinctive reactions and from ordinary conditioning. To appreciate this difficulty will require a brief digression.

When a breach is made in a termite nest, the workers set up a distinctive pounding upon the floor of the tunnel and the warriors come charging to the spot, where they take up defensive positions. Is this pounding sound a sign to the warrior-termites that there is a breach in the nest? It happens that this behavior is purely instinctive, and most students of sign-behavior believe that signs must achieve their signification through *learning*. But is learning a sufficient criterion? Are all stimuli that elicit learned reactions automatically signs? In developing any skill, such as tying the shoes, the proprioceptive stimuli produced by one response become conditioned to the succeeding response—but of what are these proprioceptive stimuli signs? With repeated experience on an electrified grill a rat will often learn to rear up on its hind legs and alternately lift them, this act apparently reducing the total intensity of pain—the painful stimulation is thus conditioned to a new response, but of what is the pain a sign?

If only some of the stimuli which elicit learned responses are signs, we must seek a reasonable distinction *within* the class of learned behaviors. We cannot draw a line between human and subhuman learning: the buzzer is operationally as much a sign of shock to the rat in avoidance-training experiments as are dark clouds a sign of rain to the professor—both stimuli elicit reactions appropriate, not to themselves, but to something other than themselves. Is voluntariness of response a criterion? Meaningful reactions may be just as involuntary as perceptions—try to observe a familiar word and avoid its meaning! Is it variability of response to the stimulus? Meaningful reactions may be just as stable and habitual as motor skills.

The Mediation Hypothesis

I shall try to show that the distinguishing condition of sign behavior is the presence or absence of *a representational mediation process* in association with the stimulus. This conception of sign behavior is based upon a general theory of learning rather than being concocted specifically to account for meaning as seen in human communication.[3] The essence of the viewpoint can be given as follows:

1. *Stimulus-objects ($\dot{S}$) elicit a complex pattern of reactions from the organism, these reactions varying in their dependence upon presence of the stimulus-object for their occurrence.* Electric shock galvanizes the rat into vigorous jumping, squeaking, and running activities, as well as autonomic "anxiety" reactions. Food objects elicit sequences of salivating, chewing, lip-smacking, and so forth. Components like salivating and "anxiety" are relatively independent of the food or shock stimulation respectively and hence can occur when such objects are not present.

2. *When stimuli other than the stimulus-object, but previously associated with it, are later presented without its support, they tend to elicit some reduced portion of the total behavior elicited by the stimulus-object.* This reduction process follows certain laws: (*a*) mediating reactions which interfere with goal-achievement tend to extinguish; (*b*) the more energy expenditure involved in making a particular reaction, the less likely it is to survive the reduction process; (*c*) there is evidence that certain reactions (e.g., autonomic) condition more readily than

---

[3] This hypothesis, as an elaboration from Hullian theory (**43**), is described in my forthcoming book, *Method and Theory in Experimental Psychology*.

others (e.g., gross skeletal) and hence are more likely to become part of the mediation process—this may merely reflect factor (*b*) above.

3. *The fraction of the total object-elicited behavior which finally constitutes the stable mediation process elicited by a sign* ($\boxed{S}$) *will tend toward a minimum set by the discriminatory capacity of the organism.* This is because the sole function of such mediating reactions in behavior is to provide a distinctive pattern of self-stimulation (cf., Hull's conception of the "pure stimulus act").

4. *The self-stimulation produced by sign-elicited mediation processes becomes conditioned in varying strengths to the initial responses in hierarchies of instrumental skill sequences.* This mediated self-stimulation is assumed to provide the "way of perceiving" signs or their "meaning," as well as mediating instrumental skill sequences—behaviors to signs which take account of the objects represented.

Whereas Morris linked sign and object through partial identity of object-produced and disposition-*produced* behaviors, we have linked sign and object through partial identity of the "disposition" *itself* with the behavior elicited by the object. Words represent things because they produce some replica of the actual behavior toward these things, as a mediation process. This is the crucial identification, the mechanism that ties particular signs to particular stimulus-objects and not to others. Stating the proposition formally: *a pattern of stimulation which is not the object is a sign of the object if it evokes in an organism a mediating reaction, this* (a) *being some fractional part of the total behavior elicited by the object and* (b) *producing distinctive self-stimulation that mediates responses which would not occur without the previous association of nonobject and object patterns of stimulation.* This definition may be cumbersome, but all the limiting conditions seem necessary. The mediation process must include some part of the same behavior made to the object if the sign is to have its particularistic representing property. What we have done here, in a sense, is to make explicit what may be implicit in Morris' term "disposition." The second stipulation (*b*) adds the learning requirement—the response of warrior-termites to pounding on the tunnel floor is ruled out since it does not depend upon prior association of pounding with discovery of a breach in the nest.

Paradigm *A* in Figure 1 gives an abbreviated symbolic account of the development of a *sign*, according to the mediation hypothesis. Take for illustration the connotative meaning of the word SPIDER. The stimulus-object ($\dot{S}$), the visual pattern of hairy-legged insect body often encountered in a threat context provided by other humans, elicits a complex pattern of behavior (R_T), which in this case includes a heavy loading of autonomic "fear" activity. Portions of this total behavior to the spider-object become conditioned to the heard word, SPIDER. With repetitions of the sign sequence, the mediation process becomes reduced to some minimally effortful and minimally interfering replica—but still includes those autonomic reactions which confer a threatening significance upon this sign. This mediating reaction (r_m) produces a distinctive pattern of self-stimulation (s_m) which may elicit a variety of overt behaviors (R_X)—shivering and saying "ugh," running out of a room where a spider is said to be lurking, and even refusing a job in the South, which is said to abound in spiders.

The vast majority of signs used in ordinary communication are what we may term *assigns*—their meanings are literally "assigned" to

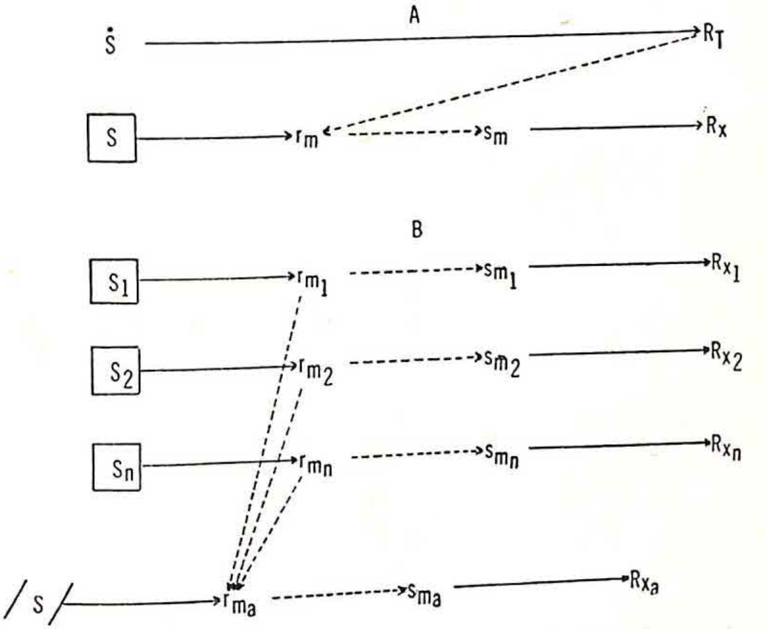


FIG. 1. SYMBOLIC ACCOUNT OF THE DEVELOPMENT OF SIGN PROCESSES: *A*. DEVELOPMENT OF A SIGN; *B*. DEVELOPMENT OF AN ASSIGN.

them via association with other signs rather than via direct association with the objects represented. The word ZEBRA is understood by most six-year-olds, yet few of them have ever encountered zebra-objects themselves. They have seen pictures of them, been told they have stripes, run like horses, and are usually found wild. As indicated in Figure 1 (*B*), this new stimulus pattern, ZEBRA, "picks up" by the mechanisms already described portions of the mediating reactions already elicited by the primary signs. In learning to read, for example, the "little black bugs" on the printed page are definitely assigns; these visual patterns are seldom directly associated with the objects signified, but rather with auditory signs (created by the child and teacher as they verbalize). Obviously, the more quickly the child can learn to make the right noises to these visual stimuli (the modern phonetic approach to reading), the more quickly these new, visual assigns will acquire significance. The child already has meanings for HOUSE, DOG, and even TYPEWRITER as *heard* stimulus patterns, but these mediation processes must be assigned to *seen* stimulus patterns.

It is apparent from the foregoing that the meanings which different individuals have for the same signs will vary with their behaviors toward the objects represented. This is because the composition of the mediation process, which *is* the meaning of a sign, is entirely dependent upon the composition of the total behavior occurring while the sign-process is being established. This indicates that to change the meaning of signs we must change behavior with respect to objects (keeping in mind that the "objects" for assigns are other signs). On the other hand, meanings are quite independent of the stimulus characteristics of the signs themselves, a point repeatedly stressed by linguists. According to the present theory, there is nothing sacred about the particular mouth-noises we use in communication any more than there is about the buzzer that becomes a sign of shock to the rat—a flash of light or a blast of air would serve as well. Of course, in human communication (in contrast to sign-behavior in the rat) it is necessary that the users of signs be able to produce as well as receive them.

EVALUATION OF EXISTING TECHNIQUES OF MEASUREMENT

Physiological Methods

Meaning has been identified with representational mediation processes. Although no conclusive evidence as to the nature of these processes can be given at this time—whether they require participation of peripheral motor and glandular mechanisms or are sufficiently characterized as central phenomena—it is convenient to conceive of them as implicit response processes which produce self-stimulation. The meager evidence available certainly does not refute this view. In any case, the investigator is encouraged to see whether or not any physiological measures display covariance with changes in meaning. Correlates of this sort would be direct indices of meaning.

Action potentials in striate musculature. Working under the impetus given by the Würzburg School of imageless thought, introspective psychologists of another generation tried to tease out the sensory content of ideas. Students like Marbe (**69**), Clark (**21**), Washburn (**110**), Comstock (**25**) and Crosland (**26**) agreed in finding kinaesthetic sensations present as a residue when everything but "meaningful thought" was excluded. But here the limitations inherent in the introspective method made themselves apparent: pressed to the limits of critical self-analysis, the trained human observer reported vague muscular and organic sensations as being present during thought—but did these sensations constitute thoughts and meanings themselves or were they merely a background of bodily tonus? The method did not permit this kind of discrimination.

Experimentalists picked up the problem at precisely the point where the introspectionists had per force dropped it—was it possible that sufficiently sensitive apparatus could record the minimal motor events that eluded conscious analysis? Watson's (**111**) statement about thought being implicit speech was the "open sesame" for a small host of gadgeteers (Wyczoikowski, **116**; Reed, **89**; Clark, **22**; Thorson, **104**) who filled subjects' mouths with an astounding variety of mechanical devices and then had them both think and mumble unusual items like "psychology." That they found little or no correspondence between the movements made during thought and speech is not particularly surprising. The "thought" movements were results of irritation in all probability, and (contrary to Watson) there is no a priori reason to expect relevant mediation processes to be restricted to the organs of speech.

Apparatus of sufficient sensitivity came with the development of electrical methods of recording and amplifying action. Electrodes placed near the motor end-plates of efferent fibers pick up minute changes in potential which cause deflections in a sensitive string galvanometer. These signals are fed through a vacuum tube amplifier, magnified thousands of times, and recorded on photographic film. Using this method, Jacobsen (**44**) and Max (**72, 73**) obtained suggestive correlations between introspectively specifiable events and objectively recorded muscular events. Jacobsen showed, for example, that when a subject, trained in techniques of progressive relaxation, *imagines* lifting his right arm, distinctive potential changes are recorded from muscles in that member but not elsewhere. Max, using deaf-mutes as an ingenious control, was able to show that these individuals display more frequent and larger potential changes in the muscles controlling their fingers while solving mental problems and while dreaming than normal individuals; he also described a negative correlation between the magnitude of such implicit activity and intelligence scores, i.e., more intelligent subjects showed less overt activity.

Are these recordable motor events the mediating reactions (r_m) in the mediation processes we have specified as essential in sign behavior? Does this technique provide an index for the presence or absence of meaning, its degree, and quality? This is an attractive possibility, but the index is a crude one. There is no way of "reading" the meaning of a sign to a subject from the recorded activity. The ultimate criterion of meaning is still introspection of the subject—he verbalizes meaning while the experimenter scurries about his periphery trying to pick it up on instruments, and when activity fails to appear at predicted locus *a* it is assumed it must be occurring at some other locus *b*. No satisfactory demonstration of the *necessity* of the motor component has been offered; meaning might be present without measurable motor activity which, when it does occur, is simply a specific overflow of excitation into motor pathways—an epiphenomenon. The same criticism applies with equal force to the other physiological correlates that have been studied.

Salivary reaction. Another pioneer investigation into the organic correlates of meaning was that by Razran (**87**), serving as his own subject. Meaningfulness of a series of signs was the independent variable, the stimuli being words for "saliva" in languages with which Razran had varying familiarity. Amount of salivary secretion was the dependent variable—following presentation of each stimulus, a dental cotton-roll was placed in his mouth for two minutes and its weight determined immediately afterward. As "meaningless" controls he used the Gaelic word for saliva, the nonsense syllables QER SUH, and periods of "blank consciousness." Salivation was greatest in his childhood tongue (Russian), next in his most proficient one (English), and less in three slightly known languages (French, Spanish, and Polish). The control conditions showed no differences among themselves, despite the fact that Razran "knew" the Gaelic word stood for saliva. This experiment demonstrates a relation between amount of salivation and degree of meaningfulness of signs to a sophisticated subject. We have here another feasible index of some aspect of meaning, albeit a very limited one.

The galvanic skin response. The GSR is one of several indices of autonomic activity, and to the extent that meanings include emotional components this measure should be useful. There are a large number of studies using the GSR that are remotely related to this problem: GSR is readily elicited by any warning or preparatory stimulus that precedes shock (Darrow and Heath, **28**; Switzer, **101**; Mowrer, **80**), and it, therefore, may serve as an indication that the preparatory stimulus has become a sign of shock. It has been used to index the intensity of pleasant or unpleasant connotations of words and experiences (Jones, **46**; Lynch, **63**). In connection with free association, GSR has been found to be a good indicator of the emotional effect of stimulus words (Jones and Wechsler, **47**). It is unfortunate, therefore, that the two most pertinent experiments in this area leave much to be desired in the way of methodological finesse.

Mason (**70**) asked this question: do changes in GSR accompany changes in meaning? What were called three types of change in meaning were studied: "certainty of meaning" (the expression of certainty by the subject as to the correctness of his recall in learning a list of nonsense syllables), "discovery of meaning" (the point in a series of readings of a trick sentence without punctuation at which the subject achieved insight), and "loss of meaning" (where the subject pressed a foot-pedal whenever the continuously vocalized word "tangerine" seemed to lose its meaning). Although the procedures and results of this experiment are reported in great detail, no tests of significance were employed and none of the necessary controls was introduced. In the first two experiments, for example, it is impossible to determine whether the deflections in GSR should be attributed to change in meaning or simply to vacillations in emotional stress (the latter seems quite likely). In the experiment on loss of meaning, no check was made to see what effect simply pressing the indicator pedal and seeing a signal light come on might have had upon GSR.

Bingham (**3**) measured psychogalvanic reaction to 72 words "selected from the educational and philosophical writings of John Dewey and Rabindranath Tagore as the most frequently occurring words in samples of 8,000 from each writer." After the galvanic measurements were made, the 50 undergraduate subjects rated the words on a three-point scale in terms of their personal meaningfulness, significance, and importance (combined into a single scale, MSI). These untrained subjects further introspected on the sensory content in the meaning of each of these words, being requested to report any visual, auditory, kinaesthetic, etc., imagery or sensations. Words having the highest MSI ratings yielded the greatest average change in skin resistance. High MSI words also had more "organic" sensory content, according to the introspections (the high "organic" content words were *intellectual*, *freedom*, *God*, *truth*, and *love*, in this order!). Much remains to be done in these directions.

Learning Methods

There are many learning studies employing meaningful materials, but rarely is meaning itself the experimental variable. Even where meaning has been deliberately varied, interest has generally centered on the effect upon learning rather than upon the use of learning as an index of meaning. Only the more relevant studies will be considered here.

Semantic generalization. When a reaction conditioned to one stimulus transfers to another, and the amount of transfer varies directly with the similarity between the two stimuli, we speak of stimulus generalization. The operations whereby semantic generalization is demonstrated are the same—except that the necessary similarities lie in meaning rather than in objective physical characteristics of the eliciting stimuli. There is no physical similarity between the *word* "blue" and light of 450 mμ yet generalization between such stimulations is easily demonstrated—we infer that the common overt reaction is mediated by some common implicit process. In the experiments to be summarized here, the precise nature of the overt reaction is unimportant—all the standard CR's have been used, salivation, GSR, finger retraction, pupillary reflex, and so on. Much of the research in this area is contributed by Russian investigators and, unfortunately, the available reports are mostly in the form of brief abstracts.

1. *From object to sign.* Kapustnik (49) set up conditioned salivary reactions to visual and auditory stimuli, transfer to verbal signs for the original cues being tested. Kotliarevsky (59) employed a cardio-vasomotor reflex: following conditioning to the sound of a bell, response to the word "bell" was tested. Metzner (75) reports a similar experiment with the pupillary reflex. In all these cases significant amounts of generalization were obtained. Traugott (107) was able to demonstrate the generalization of conditioned inhibition from blue light to the words, "blue, blue"—in fact, these words (quite different as physical stimuli) showed greater generalization than actual red light. This investigator was also able to show that the inhibitory effect of the words "blue, blue" transferred to *other* conditioned reactions more broadly than did the effect of blue light itself, a finding which fits well with our notions as to the abstractability of sign-processes. Traugott and Fadeyeva (108) combined conditioning and free association techniques: with excitatory CR's set up to a bell-whistle-light pattern and inhibitory CR's to a whistle-touch pattern, free associations to the verbal signs of these stimuli were recorded along with the latencies of these associations. Association to words representing conditioned excitors were made more rapidly than those to words representing conditioned inhibitors and, interestingly enough, after extinction of the excitatory CR's associations became slower and generically older, i.e., associations which had referred to the experimental situation now referred to pre-experimental situations.

2. *From sign to sign.* When a response is conditioned to one sign (e.g., the word TREE) and generalization to other signs (e.g., BUSH or the picture of a tree or bush) is measured, the essential role of meaningful mediation is merely more obvious than in the preceding situation. Razran (88) flashed single words and short sentences on a screen while six adults were eating—conditioned salivation developed rapidly. In a second session, different words and sentences were used to measure generalization. For the single words generalization was found to be greater for semantically related words (e.g., STYLE to FASHION) than for phonetically related words (e.g., STYLE to STILE), a result in keeping with Traugott's results discussed above and data more recently obtained by Riess (90) using the GSR as a measure. A slightly discordant note is contributed by Keller (55). After conditioning the GSR to a *picture* of a boy-scout hat, tests for generalization were made for a picture of a fireman's hat and the printed word HAT, neutral control items being pictures of a duck and baseball and the words DUCK and BALL. While significant generalization to the picture of a fireman's hat occurred, no transfer to the word HAT was obtained. Keller argues reasonably that if the generalization between the two pictured hats was based on a common mediating response, thinking or subvocalizing "hat," the printed word should have shown the same effect. In one of the most interesting studies of this type, Riess (91) has related semantic generalization to stages in genetic development. Four groups of subjects, varying in mean age from 7:9 years to 18:6 years of age, were trained to give the GSR to selected verbal stimuli. Tests for transfer to synonyms, antonyms, and homonyms of the original words were run. The generalization results indicate that meaningful or semantic similarities (synonym and antonym relations) increase in importance as the individual matures while the importance of physical similarities (homonym relation) decreases.

3. There are other semantic relations that could be studied with similar

techniques, but there is little evidence available. Generalization between *hierarchial levels of signs*, e.g., between DOG and ANIMAL, has been studied by Goodwin, Long, and Welch (**36**), but there are certain difficulties with the research design (see below). Generalization would also be expected to occur *from sign to object*, even though this seems to reverse the sequence followed in the development of meaning. Kapustnik (**49**) found that salivary reactions conditioned to verbal signs transfer to the stimuli signified. This is the only directly relevant study I have uncovered, but observations on much of social behavior fit the paradigm. The prejudicial reactions associated with "Wop" and "Jap" on a verbal stereotype level certainly tend to transfer to the social objects represented, once they are encountered. Finally, mention should be made of generalization *from object to object via semantic mediation*. The reverential care with which the adolescent handles a certain handkerchief, a certain lock of hair, a certain lipstick-printed napkin has nothing to do with physical similarities among these objects themselves. Similarly, an "inferiority complex" may render a wide range of physically dissimilar social objects and situations equivalent in meaning and hence reaction to a particular individual (cf., G. W. Allport's [**1**] trait hypothesis).

Cofer and Foley have related the various studies of semantic generalization to the theoretical mediation process. They state that semantic generalization "thus presupposes and depends upon the preexperimental formation of conditioned responses or associations, i.e., *the gradient of generalization is a gradient along a dimension of conditioned stimulus functions*. The stimuli need be similar only in so far as they have previously been conditioned to the same (or similar) response" (**23**, p. 520). The pre-experimental formation of conditioned responses to which they refer is a special case of the formation of mediation processes, as discussed earlier in this paper. It is probably necessary to assume that primary generalization occurs both among mediating reactions and the stimuli they produce, in order to account for the fact that *gradients* of semantic generalization are correlated with *degrees* of meaningful similarity.

Transfer and interference in learning. The experiments contributed by Cofer and Foley in support of their hypotheses fit the standard transfer design. The general procedure was as follows: first a single repetition on a *buffer list* of numbers (spelled out) was given as warm-up; then an *equating list* of proper names was presented once and scored for recall immediately, subjects being assigned to various experimental conditions on this basis; each subject was then given four unscored repetitions on either a *reinforcement list* or a *control list;* finally, all subjects were tested for recall of a *test list* of words immediately after a single presentation. Experimental, control and test lists used in the first experiment (Foley and Cofer, **32**) are given in Table 1. All words on a given reinforcement list bear the same relation to the test list, either some degree of synonymity or some degree of homonymity.

Although all experimental conditions yielded better recall scores than the control condition, there are several curious points about these results. In the first place, the best homonym list shows more "generalization" to the test list than the best synonym list—this is in flat contradiction to most other findings (cf., Traugott, Razran, Riess above). Secondly, the difference in recall between the two homonym lists is clearly greater than that between the two synonym lists, yet the former are obviously equivalent (both being identical in sound to the test words) while the latter are definitely not equivalent (the words on I are close synonyms of the test words but those on II bear no relation whatever to the test words, viz., sent-killed, vein-help, pear-result, sow-factory). Fortunately, the difference between "Synonym II" and the control condition was not significant.

TABLE 1

MATERIALS FOR AN EXPERIMENT ON MEDIATED TRANSFER (FOLEY AND COFER, 32)

Control List	*Reinforcement Lists*				*Test List*
	Homonym I	*Homonym II*	*Synonym I*	*Synonym II*	
palm	cent	scent	dispatched	killed	sent
set	vain	vane	vessel	ship	vein
reed	pare	pair	fruit	result	pear
very	sew	so	plant	factory	sōw
numb	rain	rein	rule	principle	reign
me	seas	seize	looks	appearance	sees
day	write	rite	just	barely	right
snap	noes	nose	apprehends	arrests	knows
rope	meat	mete	join	enlist	meet
spire	dō	doe	batter	bruise	dough
	Mean Number of Words on Test List Recalled Following Above				
4.80	6.72	5.64	5.88	5.24	

In searching for an explanation of these points, a flaw in design was discovered which renders this entire technique suspect. Since all the words on a given reinforcement list bore the same relation to those on the test list, all a subject had to do was to "catch on" to this abstract relation and then proceed to manufacture the test list rather than recall it. Given four trials on a list of only 10 meaningful words, the subject presumably masters most of it. If, on the single presentation of the test list, he now notes that the new words are homonyms of the old, by merely recalling *cent*, *vain*, and *pare* he can do a pretty good job of manufacturing (and checking by recognition) *sent*, *vein*, and *pear*. He can do the same thing with synonym lists, but here he will make more errors since there are more alternatives. The same loophole in design is even more apparent in later studies in this series. In a study on *antonym gradients*, for example, Cofer, Janis, and Rowell (24) themselves point out that 19 of 28 subjects reported that they recognized the opposition relation. Foley and Mathews (34) and Goodwin, Long and

Welch (36) report experiments using the same method, and their results are of dubious value for the same reasons.

Two experiments on interference in verbal learning by Osgood (83, 84) were explicitly designed to test certain hypotheses regarding the nature of meaning. The following hypotheses were set up: (1) Words of opposed meaning are so because they elicit reciprocally antagonistic mediating processes. (2) Repeated reinforcements of the association between a new stimulus and a particular mediating reaction produce a negatively accelerated increase in the excitatory tendency associating this stimulus with this reaction *and simultaneously an equal inhibitory tendency associating this stimulus with the reciprocally antagonistic reaction*. In other words, in learning to make a reaction to a stimulus the organism is simultaneously learning *not* to make the directly antagonistic reaction to that stimulus. (3) Both excitatory and inhibitory tendencies generalize in the usual fashion among both mediating reactions and the stimuli they produce.

With nonsense letter-pairs as constant stimuli and meaningful adjectives as varied responses in the standard retroactive interference paradigm (*A–B*; *A–K*; *A–B*), it was predicted that for both transfer and retroaction tests interpolated responses *similar* in meaning to the original responses should show the least interference (generalization of excitatory tendency) and responses *opposed* in meaning should show the most interference (generalization of inhibitory tendency), as compared with an intermediary neutral condition. The total design was such that each subject learned an equal number of items in each meaningful relation, thus avoiding the type of set that troubled the Foley and Cofer studies. The results for both transfer and retroaction situations were essentially those predicted (83). The second study (84) offered further evidence that a special form of reciprocal inhibition is operating in the successive learning of opposed meanings for the same sign. Different groups of subjects were given varying degrees of learning on the interpolated materials (cf. the design used by Melton and Irwin, 74) and only similar and opposed meaningful responses were compared. The learning of opposed responses was characterized by longer latencies of reaction and more frequent blanks (failures of response), both increasing with the degree of interpolated learning. Since both these phenomena are characteristic of weakened habit strength, they follow from the hypotheses.

These findings point to the following general conclusion: *when a sign or assign is conditioned to a mediator, it will also tend to elicit other mediators in proportion to their similarity to the original reaction; it will tend to inhibit other mediators in proportion to the directness of their antagonism to the original reaction.* In everyday language, this indicates that signs which develop a certain meaning through direct training will readily elicit similar meanings but resist being associated with opposed meanings. If the sign RUSSIAN means *bad* to the conservative college student he easily accepts substitution of *dirty*, *unfair*, and *cruel*, but it is difficult for him to think of Russians as *clean*, *fair*, and *kind* (cf., Stagner and Osgood, **100**).

Perception Methods

There is an intimate relation between perceptual and meaningful phenomena (**14**, **105**). It is borne out by the confusions psychologists display in using these terms. In one of Maier's (**67**) ingenious insight situations, for example, the crux of the problem lay in whether or not the human subject could shift from utilizing the handle of an ordinary lab clamp as something to tighten (original use) to something to hang one's hat on (use which would solve the problem). We could say that this handle must be "perceived differently" or the "field restructured perceptually" (cf., Köhler, **58**; Wertheimer, **112**), or that it must be "given a new functional value" (cf., Duncker, **30**), or that it must "acquire a new meaning or significance as a stimulus." The voluminous literature on memory for forms has been interpreted both as demonstrating perceptual dynamics (cf., Koffka, **57**) and semantic dynamics (cf., Bartlett, **2**)—witness particularly the experiment by by Carmichael, Hogan, and Walter (**15**) in which the deliberate introduction of different meaningful words in association with the same abstract forms markedly influenced the way they were recalled.

As was the case with the learning approach, there are few experiments in which meaning has been deliberately introduced as a variable. Remotely relevant are a group of studies in which the effect of hunger upon perception in ambiguous situations has been measured (Sanford, **93**, **94**; Levine, Chein, and Murphy, **62**; McClelland and Atkinson, **64**). The way of perceiving the ambiguous stimuli was clearly modulated by the presence or absence of this motive state. Postman and Bruner (**86**) have studied the effect of a different motive state, frustration, upon the perception of tachistoscopically presented sentences. Most significant from our present point of reference was the marked increase in "aggressive" and "escape" words as misperceptions following frustration. Generalizing, we might say that the internal *motive state* of the individual, as part of the total stimulus context, changes the probabilities of occurrence (availability) of alternate mediating processes for the same external stimulus. Whether the known *value* of an object, as one dimension of its meaning, can influence the way it is perceived is a moot question at this time. Bruner and Goodman (**12**), with apparent size of coins as the perceptual characteristic measured and rich or poor 10-year-olds as subjects, obtained what may be interpreted as positive results; Carter and Schooler (**18**), under generally similar conditions, failed to substantiate the earlier conclusions. Mausner and Siegel (**71**), using recognition-time as a measure and stamps of varying value as stimuli, also report negative conclusions. Most relevant to the problem of measuring meaning are the following experiments.

Bruner and Postman (**13**) compared the apparent sizes of a dollar sign (positive symbol), swastika (negative symbol), and an abstract geometrical design (neutral control), as estimated by manipulating a spot of light when plastic discs of identical size bearing these symbols were held in the subject's hands. Both dollar sign and swastika showed significant *over*estimation. According to the investigators, two dynamic processes were operative: (a) perceptual enhancement due to the positive value of the dollar sign and (b) perceptual accentuation of apparent size due to the swastika alerting the organism to danger or threat—a single process of enhancement in size due to distinctiveness might be more parsimonious, and there are many other possible hypotheses.

Postman, Bruner, and McGinnies (**86**) hypothesized that personal values (as defined by scores on the Allport-Vernon test) are among the behavioral or attitudinal determinants of perception. Twenty-five subjects were shown 36 words, one at a time, in a modified Dodge tachistoscope, these words being presented in random orders and being chosen to represent the six Allport-Vernon values. The usual method for obtaining recognition thresholds was used—gradually increasing the flash exposure time until the subject correctly identifies the word. Take for illustration a subject with high social values and low theoretical: according to the investigators, his threshold for perceiving words like "loving" and "devoted" should be lowered by *selective sensitization;* in the presolution period he should misperceive words covaluant with the correct word because of *value resonance;* and his threshold for words like "verify" and "research" should be raised because of *perceptual defense.* The results were consistent with the general thesis, but the single principle of selective sensitization seems sufficient to explain them.

McGinnies (**65**) inquired more penetratingly into the matter of "perceptual defense," conjecturing that autonomic reactions are aroused prior to conscious awareness of the meaning of a threatening word and hinder its perception. A list of 11 neutral words and seven emotionally charged words were presented tachistoscopically and recognition thresholds determined in the usual manner; GSR was recorded as a measure of emotional disturbance. Taboo words were found to require longer exposures for recognition and their prerecognition presentations were accompanied by significantly stronger emotional reactions. When asked if they had reported their perceptions promptly and accurately, all undergraduate subjects said they had.

Both this study and the preceding one on values have come in for their share of criticism. Howes and Solomon (**40**) argue that it is unnecessary to appeal to "selective sensitization" in the former case and "perceptual defense" in the latter, since recognition thresholds for words have been shown to vary with their frequency of usage or familiarity (Howes and Solomon, **41**) and this would provide a parsimonious explanation of both sets of findings. In the case of Allport-Vernon value systems, it seems reasonable to suppose that people with high theoretical values (and hence presumably with more courses, books, etc., in scientific fields) will have had more frequent visual contact with words like "logical" and "research"; in the case of taboo words, it also seems reasonable to suppose that the frequency of visual contact with words like "whore," "penis," and "bitch" is much less than with such control words as "child," "clear," and "dance." McGinnies' argument that these taboo words are much more frequent in ordinary conversations seems to be largely beside the point, since the test conditions were visual. The other main point raised by Howes and Solomon is that the emotional reactions *accompanied* recognition of the taboo words, appearing to precede simply because subjects inhibited reporting them—particularly since a member of the opposite sex was always present. These critics draw a delightful picture of what might have been going on in the subjects' minds during the McGinnies experiment. In an answer to these criticisms, McGinnies (**66**) draws a different picture. This is another issue yet to be resolved by further research. In this connection, a recent study by Lazarus and McCleary (**61**) reports that subjects may show heightened GSR to nonsense syllables previously associated with shock even when exposures are stopped prior to actual recognition. But here again one questions the validity of the demonstration—how can one react emotionally to the meaning of a sign before its significance has been appreciated? It may be that we will be forced to accept some conception of "unconscious" and "conscious" levels of perception or meaning.

Skinner (**95**) devised a "verbal summator" technique for studying language behavior which resembles these perception methods. Samples of meaningless speech sounds, obtained by permuting and combining elemental phonemes—a sort of verbal inkblot—are repeated until the subject himself perceives some meaningful form. According to Skinner, the verbal summator "evokes latent verbal responses through summation with imitative responses to skeletal samples of speech." That this method gets at the comparative strengths of verbal habits is indicated by the fact that the same double-log function of frequency to rank (Zipf's Law) appears when a large sample of such responses are analyzed. Estes (**31**) has described a visual form of summator which presents skeletonized verbal materials tachistoscopically.

Association Methods

Freud would have been the first to point out that the associations produced when a patient "allows one idea to lead to another" are in no sense "free" or random, but rather are semantically determined. Another analyst, Jung (48) used a more formal association method to get at the meanings of words to individuals. Lists of verbal stimuli calculated to touch off complexes were imbedded among neutral words. Among the indices of "unusual" responses was the rareness of the association itself. In order to judge the commonness or rareness of particular associations, it was necessary to know the comparative frequencies with which various responses to a given stimulus word occur in a representative sample of the population. Kent and Rosanoff (56) obtained responses to 100 common English nouns and adjectives from 1,000 subjects; their sample must have been fairly representative because the occasional re-checks that have been made show rather surprising agreement. Given norms like these, the unusualness of a subject's responses can be indexed by the frequency of occurrence of that response in the populations; *sharp* can be expected as a response to NEEDLE 152 times per 1,000 (15.2%), but *weapon* occurs only once (00.1%).

The gross majority of word associations are semantically determined, i.e., result from the mediation process set in motion by the verbal stimulus as a sign. All such associates are similar in some way to the stimulus word, either similar in meaning (NEEDLE–*pin*), which would include hierarchial relations (NEEDLE–*tool*), or in terms of commonness of context (NEEDLE–*thread*). The venerable associationistic principles of similarity and contiguity will be recognized here. After an intensive analysis of ways of classifying associates, Karwoski and Berthold (50) conclude that nearly all responses can be categorized as either some form of similarity or contrast. What about contrast responses? The single most frequent associate is often the direct opposite (LIGHT–*dark;* MAN–*woman*). For a number of reasons, this writer believes such contrast responses are *not* semantically determined at all, but rather reflect overlearning of verbal skill sequences, quite akin to FOOT–*ball*, APPLE–*cart*, and WASTE–*basket*. The tendency to free associate opposites increases with age, children readily giving similar and contextual responses but rarely opposites (cf., Woodworth, 114, p. 346). Furthermore, rather than being distributed among many varied but roughly equivalent words, as is the case with similar associates, the opposition tendency is largely restricted to a single word, the direct opposite (cf., Kent and Rosanoff tables). Karwoski and Schachter (54) report this same effect with opposites and add the fact that opposites are given with significantly shorter reaction-times than similars.

One of the more interesting applications of the association method has been in differentiating responses to sign, symbol, and object levels of stimulation. In an early study on this, Dorcus (29) compared associations to color words (signs) and actual bits of colored paper (objects).

Whereas co-ordinate and contrast responses were most common to color signs (WHITE–*black*, 'RED–*blue*), the names of contextually related objects were most commonly given to color objects (BLUE PAPER–*ribbons;* RED PAPER–*fingernails*). More recently Karwoski, Gramlich, and Arnott (**51**) have obtained associations to visually perceived actual objects, pictures of these objects, and verbal labels for these objects. The stimulus materials were such everyday things as *pipe*, *leaf*, *dollar*, and *pistol*. Where differences appeared, the dividing line was typically between the verbal level and the other two modes. For example, the most common response to the *word* FORK was, of course, *knife*—on both picture and objects levels the most common response was *eat*. Reaction times for the verbal level were also shorter.

Related to associational procedures are the effects of *context* upon meaning. It is a matter of common observation that a man's moods, emotions, and motives influence the character of his verbalizations. Bousfield and Barry (**9**) and Bousfield (**8**) found that subjects' rated moods (on a scale from "feeling well as possible" to "feeling as badly as possible") correlated with their rates of production of pleasant vs. unpleasant associates. The relatively stable attitudes of an individual also exert a contextual effect upon associations. Foley and MacMillan (**33**) have shown that associates to 40 ambiguous words (like *binding*, *administer*, *discharge*) are clearly influenced by the occupational status of subjects, as law students, medical students, or nonprofessional students. Perhaps because of the obviousness of the matter, no research seems to have been done upon the effect of the external, situational context upon meaning. This context includes the facial expressions and gestures of speakers, the objects present, and the activities underway, and so on (cf., Malinowski's [**68**] enlightening discussion of this in relation to decoding the language of another culture). Many slips of the tongue completely escape notice, simply because the situational context "carries" the intended meaning.

Howes and Osgood[4] have given attention to the manner in which the meaning of a particular sign is affected by the pattern of verbal materials within which it is imbedded. A sequence of four spoken words made up each item, the first three serving as the context and the fourth, spoken with greater emphasis, serving as the actual stimulus for word-association by the subjects. One experiment was designed to get at the effect of varying the *density* of contextual items having a common emantic direction: Group A heard three contextual stimuli of very imilar meaning (e.g., *sinister*, *devil*, *evil*–DARK), Group B had one ieutral word added (*eat*, *devil*, *evil*–DARK), Group C two neutral words *eat*, *basic*, *evil*–DARK), and Group D, as control, had all neutral words n the context. When the frequencies of response-words related to the articular context (e.g., *thief*, *mystery*, *dead*, etc.) were plotted as a unction of the density of influence in the context, number of influenced ssociates turned out to be a simple multiple of the number of relevant vords in the context. In a second experiment, three influencing words f relatively independent meaning were used as "context," and the uestion was how *temporal proximity* of contextual stimuli affects the neaning of the eliciting sign. With an item like *feminine*, *strong*, *young*–MAN, the responses clearly relevant to each contextual stimulus could e isolated (*woman*, *girl* vs. *hard*, *work* vs. *boy*, *child*, for examples). Then the frequency of occurrence of such related responses was plotted s a function of the order of presentation of their contextual stimuli e.g., frequencies of *woman* as a response when "feminine" is in third osition, nearest MAN, second position, and first position, most remote rom MAN). The results indicate that degree of influence of a con- extual stimulus upon the meaning of a sign is a sharply negatively ccelerated function of the temporal interval between them. In other vords, the influence of one word upon another falls off rapidly as the mount of intervening material increases.

---

[4] Howes, D. H., & Osgood, C. E. Studies on the combination of associative probabilities in linguistic contexts (in preparation).

Scaling Methods

Considering the number of traits, abilities, and attitudes that psy- hologists have attempted to measure by scaling methods, it is signifi- ant that there has been practically no attempt to measure meaning his way. Since many psychologists must have thought about the prob- em at one time or another, this probably reflects the general belief that neanings are too complicated or too unique, or both. The few timid teps that have been taken in this direction involved drastic limitations n the scope of measurement, being aimed at scaling one or two iso- ated dimensions of meaning rather than meaning-in-general.

One group interested in scaling meaning has been the researchers in uman learning, who wanted to be able to select materials for their xperiments which could be specified with respect to this variable. A umber of studies have been reported on the meaningfulness or "associa- ion value" of nonsense syllables (Glaze, **35**; Hull, **42**; Witmer, **113**). The typical method was to use nonsense syllables as stimuli for word-associations, "meaningfulness" being indexed by proportions of sub- ects who could find any associations. One could then select equated ists for learning experiments, equated on this one basis at least.

Also motivated to provide learning experimenters with standardized materials, Haagen (**37**) scaled 400 pairs of common adjectives in terms of their synonymity, vividness, familiarity, and association value. The method used was to have 280 college undergraduates judge these words on defined scales: (a) *synonymity* of a given word was judged on a seven-point scale in terms of the degree to which it denoted the same actions, objects or conditions as a standard word; (b) *vividness*, also judged on a seven-point scale, was defined as the clarity of graphicness of the impressions which a given word aroused; (c) *familiarity* was judged on a five-point scale, defined as the degree to which the judge knew the meaning of the word; and (d) *association value*, judged on a seven-point scale, referred to the degree to which the given word and a standard were associated in thought (e.g., hungry-thirsty, big-large, would have high association value). Useful though synonymity and associative value may be for purposes of learning experiments, they do not offer anything in the way of a measure of meaning—these judgments were always relative to some particular standard word, varying from one set of test words to another. The familiarity measure has nothing to do with meaning, of course. The vividness scale, being applied to each word separately rather than comparatively, probably is tapping some generalizable dimension of meaning.

Mosier (**79**) made the most direct application of scaling methods to the study of meaning. College subjects rated some 296 adjectives on an 11-point scale in terms of their favorableness-unfavorableness, these adjectives being selected from Thorndike's word lists as words expressing some degree of general evaluation. Frequency distributions of the responses to each word were scaled according to the method of successive intervals and plotted on probability paper. Plots for approximately 200 of these words were linear, indicating normal distribution of the data, when treated in this manner. Most of the words showed the "precipice effect" at one side or the other of the midpoint of the scale, indicating a higher degree of agreement on the *direction* (favorable or unfavorable) of the evaluation than on the *intensity*. Mosier was able to demonstrate a reasonable ordering of evaluative words in terms of their mean locations (e.g., excellent, good, common, fair, poor, etc.), including such information as the fact that "better" is connotatively less favorable than "good" (grammarians to the contrary). The most significant point is that Mosier demonstrated the feasibility of scaling certain aspects of meaning.

Summary on Existing Methods

The purpose of the preceding review has been to see if there already exist adequate methods of measuring meaning. By "adequate" I mean already meeting most of the criteria of satisfactory measuring instruments. What are these criteria? (*a*) *Objectivity.* The method should yield quantitative and verifiable (reproducible) data. (*b*) *Reliability.* It should yield the same values within acceptable margins of error, when the same conditions are duplicated. (*c*) *Validity.* The data obtained should be demonstrably covariant with those obtained with some other, independent index of meaning. (*d*) *Sensitivity.* The method should yield differentiations commensurate with the natural units of the material studied, i.e., should be able to reflect as fine distinctions in meaning as are typically made in communicating. (*e*) *Comparability.* The method should be applicable to a wide range of phenomena in the field, making possible comparisons among different individuals and groups, among different concepts, and so on. (*d*) *Utility.* It should yield information relevant to contemporary theoretical and practical issues in an efficient manner, i.e., it should not be so cumbersome and laborious as to prohibit collection of data at a reasonable rate. While this is not an exhaustive listing of criteria, it is sufficient for our purposes.

1. The *physiological measures* (including action potential, GSR, and salivary records) are of somewhat dubious validity, since there has been no demonstration of the necessity of these peripheral components, and they are not sensitive measures in that we are unable to interpret details of the records in our present ignorance. Their chief drawback, however, is cumbersomeness—the subject has to be "rigged up" in considerable gadgetry to make such measurements. For this reason, even should validity and sensitivity problems be met satisfactorily, it seems likely that physiological indices will be mainly useful as criteria against which to evaluate more practicable techniques.

2. *Learning measures* (including semantic generalization and transfer/interference methods) are also somewhat cumbersome procedurally, but their main drawback as general measures of meaning is their lack of comparability. Any measure of generalization or interference is made with respect to the original learning of some standard which necessarily varies from case to case. The chief usefulness of learning measures, therefore, lies in the test of specific hypotheses.

3. The chief drawback with *perception measures* (e.g., what is perceived in ambiguous stimulus forms, the recognition-times for tachistoscopically presented words) is that they are not valid measures of *meaning*. They get at the availability or comparative habit strengths of alternate meanings or ways of perceiving. The fact that a religious person perceives VESPERS with a shorter presentation time than a theoretically oriented person says nothing about *how* the meaning of this term differs for them; the fact that the religious person perceives VESPERS more quickly than THEORY says nothing about the difference in meaning of these two words to this individual. The same statements apply to Skinner's "verbal summator" technique.

4. The selection of responses in *association methods* is partly dependent upon the meaning of the stimulus items (and hence indexes meaning) and partly dependent upon habit strength factors. The chief drawback, as a general measure of meaning, is lack of comparability. The responses of two individuals to the same stimulus, or of the same individual to two stimulus words, are

essentially unique as bits of data. Comparability can be obtained with group data, but this limits the method.

5. *Scaling methods* can be viewed as forms of controlled association in which the nature of the association is specified by definition of the scales (favorable-unfavorable, vividness, etc.) but the direction and intensity of association is unspecified. By the very nature of the scaling method, the comparability criterion is usually satisfied (provided the subjects can be shown to agree upon the meaning of the scale and its divisions). As used by Mosier, however, the method can have only partial validity. This is because he tapped only one dimension of meaning, the admittedly important evaluative dimension, whereas we know that meanings vary multidimensionally.

THE SEMANTIC DIFFERENTIAL

The method to be proposed here is a combination of associational and scaling procedures. It is an indirect method in the same sense that an intelligence test, while providing objective and useful information, does not directly measure this capacity. However, unlike the intelligence test which treats this ability *as if* it were distributed along a single continuum (e.g., IQ scores vary along a single scale), we accept at the outset that meanings vary in some unknown number of dimensions and frame our methodology accordingly.

Research Origins of the Method

This method had its origins in research on synesthesia, defined by Warren in his *Dictionary of Psychology* (**109**) as "a phenomenon characterizing the experiences of certain individuals, in which certain sensations belonging to one sense or mode attach to certain sensations of another group and appear regularly whenever a stimulus of the latter type occurs." This implies a sort of "neural short-circuiting" that is present in only a few freak individuals, and it is true that many of the classic case histories in this area gave credence to this view: a subject reported pressure sensations about his teeth and cheeks whenever cold spots on his arms were stimulated (Dallenbach, **27**); a girl displayed a rigid system of relations between specific notes on the musical scale and specific color experiences, consistent when tested over a period of seven and one-half years (Langfeld, **60**). But here, on the other hand, was a man who imagined the number "1" to be yellow, "2" to be blue, "3" to be red . . . and, of course, "8" to be black (anyone who has played pool will recognize the origin of this system); and here was a little girl who recalled her friends as having pink faces and her enemies as having purple faces. What modalities are crossed in these cases?

A more recent series of investigations by Karwoski, Odbert, and their associates related synesthesia to thinking and language in general (cf., also Wheeler and Cutsforth, **115**). Rather than being a rare phenomenon, Karwoski and Odbert (**52**) report that as many as 13 per cent of Dartmouth College students regularly indulged in color-music synesthesia, often as a means of enriching their enjoyment of music. These photistic visualizers varied among themselves as to the modes of translation employed and the vividness of their experiences, and their difference from the general population appeared to be one of degree rather than kind. Whereas fast, exciting music might be pictured by the synesthete as sharply etched, bright red forms, his less imaginative brethren would merely agree that terms like "red-hot," "bright," and "fiery," as verbal metaphors, adequately described the music; a slow and melancholic selection might be visualized as heavy, slow-moving "blobs" of sombre hue and described verbally as "heavy," "blue," and "dark." The relation of this phenomenon to ordinary verbal metaphor is evident: a happy man is said to feel "high," a sad man feels "low"; the pianist travels "up" and "down" the scale from treble to bass; souls travel "up" to the good place and "down" to the bad place; hope is "white" and despair is "black." The process of metaphor in language as well as in color-music synesthesia can be described as the parallel alignment of two or more dimensions of experience, defined verbally by pairs of polar opposites, with translations occurring between equivalent portions of the continua (Karwoski, Odbert, and Osgood, **53**, pp. 212–221).

Interrelationships among color, mood, and musical experiences were studied more analytically by Odbert, Karwoski, and Eckerson (**81**). Subjects first listened to 10 short excerpts from classical scores and indicated their dominant moods by checking descriptive adjectives arranged in a mood circle (cf., Hevner, **39**). Then, on a second hearing, they listed the colors appropriate to each score. Significant relations were shown; the color associations to musical scores followed the moods created. A portion of Delius' *On Hearing the First Cuckoo in Spring* was judged leisurely in mood and preponderantly green in color; a portion of Wagner's *Rienzi Overture* was judged exciting or vigorous in mood and preponderantly red in color. When another group of subjects was merely shown the mood adjectives (with no musical stimulation) and asked to select appropriate colors, even *more* consistent relations appeared, suggesting that the unique characteristics of the musical selections had, if anything, somewhat obscured the purely verbal or metaphorical relations between colors and moods. Almost identical findings have been reported by Ross (**92**) for relationships between the colors used in stage lighting and reported moods produced in the audience. Data are also available for the effects of color upon mood in mental institutions and in industrial plants.

regularly *up* and *light* (*white*) in relation to *bad* things, which were *down* and *dark* (*black*). A prevalent myth tells of how the gods helped the original man to struggle "up" from the "dark," "cold," "wet," "sad" world below the ground to the "light," "warm," "dry," "happy" world on the surface of the earth. Among certain Siberian Aborigines, members of a privileged clan call themselves the "white" bones in contrast to all others who are referred to as "black" bones. And even among the Uganda Negroes we find some evidence for a white god at the apex of the hierarchy, and white cloth is clearly associated with purity, being used to ward off evil spirits and disease. Such data suggest the existence of a pervasive semantic frame of reference. Further study of the problem by more adequately trained investigators could be richly rewarding.

Stagner and Osgood (**100**) adapted this method and the logic underlying it to the study of social stereotypes. The notion of a continuum between the polar terms was made explicit by using such terms to define the ends of 7-step scales. Rather than studying the relations between continua, as above, a set of scales was used to measure the "meaning" of particular concepts, such as PACIFIST, RUSSIAN, DICTATOR, and NEUTRALITY. Successive samples of subjects were tested between April, 1940, and March, 1942 (including a sample obtained just prior to the Pearl Harbor incident). A single item on the tests appeared as follows:

PACIFIST: Kind ____:____:____:____:____:____:____ cruel

with the subject instructed to check that position on the scale which best represented the direction and intensity of his judgment. The concepts and scales related in successive items of the test were randomized to insure as much independence of judgment as possible. The feasibility and efficiency of using this method to record the changing structures of social stereotypes (e.g., the changing meanings of a set of social signs) were demonstrated. That a total shift from an essentially pacifistic to an essentially militaristic frame of reference had been accomplished, even before the Pearl Harbor incident provided the spark to overt expression, was clearly evident in the data.

More important from the point of view of methodology was the following observation: As used by our subjects in making their judgments, the various descriptive scales fell into highly intercorrelated clusters. Fair-unfair, high-low, kind-cruel, valuable-worthless, Christian-anti-Christian, and honest-dishonest were all found to correlate together .90 or better. This cluster represented, we assumed, a single, general factor in social judgments, the evaluative (good-bad) dimension of the frame of reference. Gradients like strong-weak, realistic-unrealistic, and happy-sad were independent of this evaluative group and pointed to the existence of other dimensions within the semantic framework. Enforced shifts in the apparent reference point of the observer (by having subjects judge the same concepts "as a German" or "as an Englishman") produced gross and appropriate changes in the evaluative dimension but did not disrupt the qualitative pattern of each stereotype—e.g., the stereotype GERMANS, when judged by students playing the role of Germans, was still seen as relatively more "strong" and "happy" (remember, this was during 1940–1942) than "noble" or "kind." This illustrates the kind of difficulty experienced when one tries to assume the point of view of another (cf., Stagner and Osgood, **99**).

Logic of the Proposed Method

The researches described above gave rise to the following hypotheses:

1. *The process of description or judgment can be conceived as the allocation of a concept to an experiential continuum, definable by a pair of polar terms.* An underlying notion in our research is that these "experiential continua" will turn out to be reflections (in language) of the sensory differentiations made possible by the human nervous system. In other words, it is assumed that discriminations in meaning, which is itself a state of awareness, cannot be any finer or involve any more variables than are made possible by the sensory nervous system (cf., Boring, *The Dimensions of Consciousness*, **6**). While failure to confirm this notion would not eliminate the proposed method as an index of meaning, its confirmation would greatly enhance the theoretical implications of this work.

2. *Many different experiential continua, or ways in which meanings vary, are essentially equivalent and hence may be represented by a single dimension.* This functional equivalence of many alternate continua was clearly evident in both the studies on synesthesia and those on the changing structure of social stereotypes. It is this fact about language and thinking that makes the development of a quantitative measuring instrument feasible. If the plethora of descriptive terms we utilize were in truth unique and independent of one another, as most philosophers of meaning seem to have assumed, then measurement would be impossible.

3. *A limited number of such continua can be used to define a semantic space within which the meaning of any concept can be specified.* From the viewpoint of experimental semantics, this both opens the possibility of measuring meaning-in-general objectively and specifies factor analysis as the basic methodology. If it can be demonstrated that a limited number of dimensions or factors are sufficient to differentiate among the meanings of randomly selected concepts, and if the technique devised satisfies the criteria of measurement stated earlier, then such a "semantic differential," as I have termed it, *is* an objective index of meaning. From the viewpoint of psychological theory, we may look upon the procedures followed in obtaining this measure as an operational definition of meaning, in the same sense that the procedures followed in obtaining the IQ score provide an operational definition of intelligence.

The operations followed in the present instance are explicit. They involve the subject's allocation of a concept within a standard system of descriptive dimensions by means of a series of independent associative judgments. The judgmental situation is designed to be maximally simple. Presented with a pair of descriptive polar terms (e.g., *rough–smooth*) and a concept (e.g., LADY), the subject merely indicates the direction of this association (e.g., LADY–*smooth*). We have developed two different methods for collecting data: In the *graphic method*, a pencil-and-paper technique which has the advantage that data can be collected from groups of subjects and hence very speedily, the subject indicates the intensity of his association by the extremeness of his checking on a 7-step scale. In the *judgment-time method*, which has the advantage that the subject cannot anticipate what concept is to be judged on a particular scale and hence cannot rationalize his reaction, intensity of association is indicated by the latency of the individual subject's choice reaction toward one or the other of the polar terms. In both methods each associative judgment of a particular concept against a particular descriptive scale constitutes one item. In successive items, concepts and dimensions are paired in deliberately rotated orders until every concept has been associated with every scale by every subject.

A Factor Analysis of Meaning

The procedures and results of this factor analysis will be described in detail elsewhere. A total of 50 descriptive scales, selected in terms of their frequency of usage, have been used in the judgment of 20 varied concepts, yielding a 1,000-item test. One hundred college students served as subjects. The graphic method was used.[5] The purpose of this factor analysis is to isolate a limited number of general dimensions of meaning having a maximal differentiating power, to try to bring some order out of semantic chaos. The larger the proportion of total variance in meaning accounted for by these factors, the more satisfactory will be the measuring instrument finally set up. A preliminary estimation of factors in the 50×50 matrix (each scale correlated with every other scale) indicates the existence of several roughly independent dimensions. An "evaluative factor" accounts for by far the largest portion of the variance. There is also evidence for a "strength factor," an "activity factor," and several others not clearly defined in this rough approxima-

[5] Apparatus for obtaining latency measurements from individual subjects has been constructed and will be standardized upon the reduced set of descriptive scales we hope to derive from this preliminary factor analysis. While this apparatus has the advantage that materials are projected from a film-strip and responses (directions and latencies) are photographed by a single-frame camera—all automatically—it is still applicable only to a single subject at a time and hence is time-consuming.

tion. Given such factors, it will be possible to select those specific scales (e.g., good-bad, strong-weak, active-passive, smooth-rough, hot-cold, etc.) which best represent them.

We have done some exploratory work on the use of the semantic differential as a practical measuring device. The two sets of profiles in Figure 3 will serve to illustrate the method. Two groups of only 20 subjects each differentiated the meaning of the adjective "polite"; Group I also differentiated "eager" while Group II also differentiated "burning" (as part of a larger study). Median judgments of the 20 subjects on each scale are plotted. There is high agreement on the dif-

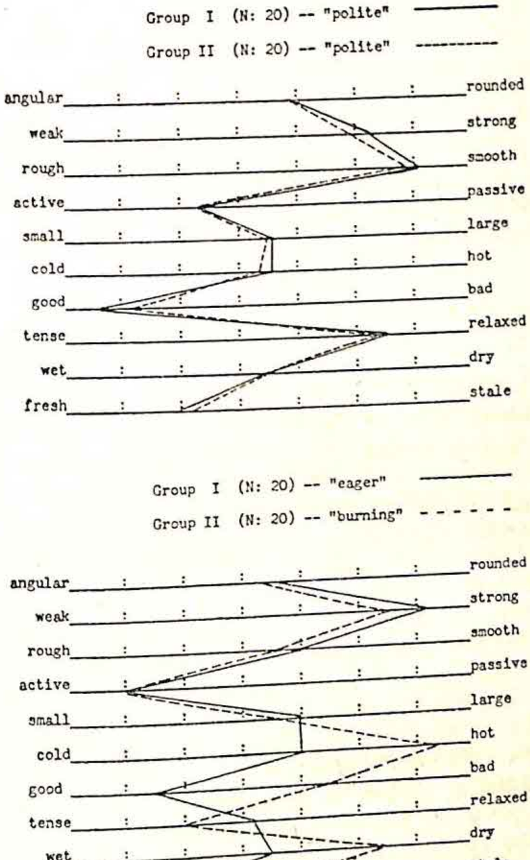


Fig. 3. Illustration of Application of a Preliminary Form of the Semantic Differential for Measuring the Connotative Meanings of Adjectives: *A*. Upper Profiles, Medians for Two Groups of 20 Subjects Differentiating "Polite"; *B*. Lower Profiles, Medians for Same Two Groups of 20 Subjects Differentiating "Eager" and "Burning."

ferentiation of the same sign, "polite." The different, but somewhat similar signs, "eager" and "burning," show significant points of discrimination: whereas they are equally *strong* and *active*, "burning" is relatively *hot* and *dry* as compared with "eager," and "eager" is relatively *good* and *fresh* as compared with "burning." These differences are obviously what they would have to be if the method has any validity. It must be emphasized that the sample of scales shown here does *not* necessarily represent those to be finally derived from our factor analysis.

Evaluation of the Method

Evaluation of this instrument against the criteria of measurement listed earlier will be the subject of future reaserch, but some evidence can be presented now.

1. *Objectivity.* The semantic differential yields quantitative data which are presumably verifiable, in the sense that other investigators can apply the same sets of scales to equivalent subjects and obtain essentially the same result.

2. *Reliability.* In the test form from which data for the factor analysis were collected, 40 of the 1,000 items were selected at random and repeated. None of the subjects was aware that this had been done. The reliability coefficient was .85. The minimum variation in profile for the two groups of 20 subjects judging "polite" is another indication of the stability of the method.

3. *Validity.* All of the data collected so far on several problems display convincing face-validity and several direct experimental checks are planned. These include (a) correlation of attitudes toward various social objects as measured on standard tests with allocation of signs of these social objects within the semantic differential, and (b) the use of experimentally induced changes in meaning of signs (cf., Stagner and Britton, **98**). We are not concerned about the problem of "labeling" factors, a point where the precision gained by the factor analytic method is often lost in the obscurities of language. Selection of specific scales to match factors can proceed on a purely objective basis, in terms of the factor loadings for each scale. As a matter of fact, the polar terms which define the scales do not admit much in the way of misinterpretation.

4. The question of *sensitivity* of the method comes down to whether it is able to reflect as fine distinctions in meaning as are ordinarily made. We have incidental evidence that a semantic differential can tease out nuances in meaning which are clearly felt but hard to verbalize deliberately.[6] If there is a real difference in the meaning of two signs, such that they would not be used in

---

[6] By way of illustration, most English-speaking Americans feel that there is a difference, somehow, between "good" and "nice" but find it difficult to explain. We gave several people these words to differentiate and it turned out that wherever "male" and "female" show a significant divergence, there also were "good" and "nice" differentiated (e.g., "good," like "male," is somewhat stronger, rougher, more angular, and larger than is "nice," which like "female" shifts toward the weak, smooth, rounded, and small directions of the space). Thus "nice man" has a slightly effeminate tone whereas "good woman" (as compared with "nice woman") has a narrowly moral tone.

precisely the same contexts, and if our measuring instrument includes a sufficient number of dimensions of the semantic space, then a significant difference should appear on at least one of the scales.

5. *Comparability.* It is here that the most serious questions arise. (a) *Is the method culture-bound?* If the tendency to dichotomize experiential continua is characteristic of Western culture but not necessarily elsewhere, then the method would not have generality. This is an empirical question requiring the skills of anthropologists and linguists for solution. (b) *Is the method limited to the differentiation of nouns against adjective scales?* The structure of our language is such that "adjectives" typically reflect abstracted qualities of experience and "nouns" the concepts and things dealt with. We have found it possible to set up scales like giant–midget, fire–iceberg, god–devil and to judge "concepts like INSINUATE and AGITATED against them. This does not seem "natural" to members of our language community, however; it is probably the stem or root meaning of words that our method taps. (c) *Can different concepts be compared?* To the extent that judgments of different concepts involve the same factor structure, any concept may be compared with any other against a single, standardized semantic framework. (d) *Can different individuals be compared?* This also comes down to the generality of the semantic factor structure. It is quite conceivable that different classes of people (scientists, ministers, etc.) have somewhat varied semantic structures, differing in the emphasis upon certain factors and interrelationships among them. In fact, a significant source of individual differences may lie here.

Our method can be criticized on the ground that it only gets at *connotative* meaning, not *denotative* meaning. This is a limitation. Both SIMON LEGREE and WAR might be allocated to approximately the same point in semantic space by our method. This would indicate similar connotative meaning, to be sure, but it would not indicate that these signs refer to the same object. Our differential will draw out the *hard, heavy, cold, ugly, threatening* connotations of the sign HAMMER, but it will not indicate that HAMMER is "an instrument for driving nails, beating metals, and the like, consisting of a head, usually of steel, fixed crosswise to a handle" (Webster's *Collegiate Dictionary*). In part, this limitation stems from our method of selecting descriptive scales in terms of frequency of usage rather than in terms of a logically exhaustive coverage, as given in Roget's *Thesaurus*, for example.

6. *Utility.* In any area of science, the development of an adequate method of measuring something (be it the wave length of radiation, blood chemistry, intelligence, or meaning) opens up well-nigh inexhaustible possibilities for application. (a) *Semantic norms.* In much the same way that Thorndike has established his norms for frequency-of-usage of common words in the English language, the semantic differential could be used to compile a functional lexicon of connotative meanings, a quantized thesaurus. Similarly, the gradual drift of changing meanings, both temporally and geographically, could be charted. (b) *Individual differences in meaning.* It is a truism that the meanings of socially significant signs differ for different classes of people. Concepts like CHURCH, LABOR LEADER, STALIN, and TRUMAN have different connotative significance to different people, and the semantic differential can be used to quantify these differences. In this sense, it can be used as a generalized, multidimensional attitude test. For example, 10 people may have identical degrees of favorableness toward NEGRO (evaluative dimension) and yet vary markedly with respect to other dimensions of the meaning-space. (c) *Changes*

in meaning. Under the pressure of events, the meanings of social signs change, e.g., the meaning of ITALIANS to Americans during the past half century. Similarly, under the "pressure" of psychotherapy, the meaning or emotional significance of certain critical concepts (e.g., FATHER, THERAPIST, ME, etc.) undergoes change. (d) *Quantification of subjective language data.* We have recently used the semantic differential as a means of scoring TAT reactions; not only is the testing process greatly speeded up, but the data are in easily manipulable form. Preliminary studies indicate that the essential individual differences in meaning of such projective materials, as teased out of complicated verbatim "stories," are sharply etched in the semantic differential data. (e) *Cross-cultural communication problems.* If the structure of the semantic space proves to be sufficiently general that the method can be translated into equivalent differentials in other languages, numerous possibilities are opened up. Are the fundamental factors in meaning and their relationships independent of the language spoken? Can the significant points of deviation in meaning of critical concepts, as between Americans and Russians, for example, be discovered? Can the finer, subtler degrees of acculturation into a new society be traced? And there are other potential applications, to aesthetics, to studying the development of meaning in children, and so on.

SUMMARY

The first portion of this paper describes a behavioral conception of the sign-process as developed from a general mediation theory of learning. The remainder is concerned with the problem of measuring meaning. Various existing approaches to the problem—physiological, learning, perception, association, and scaling methods—have been evaluated against the usual criteria of measurement and have been found inadequate. The development of a semantic differential as a general method of measuring meaning is described. It involves (a) the use of factor analysis to determine the number and nature of factors entering into semantic description and judgment, and (b) the selection of a set of specific scales corresponding to these factors which can be standardized as a measure of meaning. Using this differential, the meaning of a particular concept to a particular individual can be specified quantitatively as a particular point in the multidimensional space defined by the instrument. Some of the possible uses of such a measuring instrument are briefly indicated.

BIBLIOGRAPHY

1. Allport, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
2. Bartlett, F. C. *Remembering.* Cambridge, England: Cambridge Univer. Press, 1932.
3. Bingham, W. E., Jr. A study of the relations which the galvanic skin response and sensory reference bear to judgments of the meaningfulness, significance, and importance of 72 words. *J. Psychol.*, 1943, **16**, 21–34.
4. Bloomfield, L. *Language.* New York: Holt, 1933.
5. Boder, D. P. The adjective-verb-

quotient: a contribution to the psychology of language. *Psychol. Rec.*, 1940, **3**, 310–343.

6. Boring, E. G. *The physical dimensions of consciousness.* New York: Appleton-Century, 1933.
7. Bousfield, W. A. An empirical study of the production of affectively toned items. *J. gen. Psychol.*, 1944, **30**, 205–215.
8. Bousfield, W. A. The relationship between mood and the production of affectively toned associates. *J. gen. Psychol.*, 1950, **42**, 67–85.
9. Bousfield, W. A., & Barry, H., Jr. Quantitative correlates of euphoria. *J. exp. Psychol.*, 1937, **21**, 218–222.
10. Bousfield, W. A., & Sedgwick, C. H. W. An analysis of sequences of restricted associative responses. *J. gen. Psychol.*, 1944, **30**, 149–165.
11. Bousfield, W. A., & Barclay, W. D. The relationship between order and frequency of occurrence of restricted associative responses. *J. exp. Psychol.*, 1950, **40**, 643–647.
12. Bruner, J. S., & Goodman, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
13. Bruner, J. S., & Postman, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.
14. Brunswik, E. Die Zugänglichkeit von Gegenständen für die Wahrnehmung. *Arch. ges. Psychol.*, 1933, **88**, 377–418.
15. Carmichael, L., Hogan, H. P., & Walter, A. A. An experimental study of the effect of language on the reproduction of visual perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.
16. Carroll, J. B. Diversity of vocabulary and the harmonic series law of word-frequency distribution. *Psychol. Rec.*, 1938, **2**, 379–386.
17. Carroll, J. B. The analysis of verbal behavior. *Psychol. Rev.*, 1944, **51**, 102–119.
18. Carter, L. F., & Schooler, K. Value, need and other factors in perception. *Psychol. Rev.*, 1949, **56**, 200–207.
19. Cason, H., & Cason, E. B. Association tendencies and learning ability. *J. exp. Psychol.*, 1925, **8**, 167–189.
20. Chotlos, J. W. Studies in language behavior: IV. A statistical and comparative analysis of individual written language samples. *Psychol. Monogr.*, 1944, **56**, No. 2 (Whole No. 255), 75–111.
21. Clark, H. M. Conscious attitudes. *Amer. J. Psychol.*, 1911, **22**, 214–249.
22. Clark, R. S. An experimental study of silent thinking. *Arch. Psychol.*, 1922, **7**, No. 48. Pp. 101.
23. Cofer, C. N., & Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior. I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
24. Cofer, C. N., Janis, M. G., & Rowell, M. M. Mediated generalization and the interpretation of verbal behavior. III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32**, 266–269.
25. Comstock, C. On the relevancy of imagery to the processes of thought. *Amer. J. Psychol.*, 1921, **32**, 196–230.
26. Crosland, H. R. A qualitative analysis of the process of forgetting. *Psychol. Monogr.*, 1921, **29**, No. 1 (Whole No. 130).
27. Dallenbach, K. M. Synaesthesis: "Pressury" cold. *Amer. J. Psychol.*, 1926, **37**, 571–577.
28. Darrow, C. W., & Heath, L. L. Reaction tendencies relating to personality. In C. P. Stone, C. W. Darrow, C. Landis, and L. L. Heath. *Studies in the dynamics of*

behavior. Chicago: Univer. Chicago Press, 1932.

29. Dorcus, R. M. Habitual word associations to colors as a possible factor in advertising. *J. appl. Psychol.*, 1932, **16**, 277–287.
30. Duncker, K. On problem-solving. (Trans. by L. S. Lees from 1935 original.) *Psychol. Monogr.*, 1945 **58**, No. 5 (Whole No. 270).
31. Estes, W. K. A visual form of the verbal summator. *Psychol. Rec.*, 1940, **4**, 174–180.
32. Foley, J. P., Jr., & Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 169–175.
33. Foley, J. P., Jr., & MacMillan, Z. L. Mediated generalization and the interpretation of verbal behavior: V. "Free association" as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299–310.
34. Foley, J. P., Jr., & Mathews, M. Mediated generalization and the interpretation of verbal behavior: IV. Experimental study of the development of interlinguistic synonym gradients. *J. exp. Psychol.*, 1943, **33**, 188–200.
35. Glaze, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
36. Goodwin, J., Long, L., & Welch, L. Generalization in memory. *J. exp. Psychol.*, 1945, **35**, 71–75.
37. Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.
38. Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.
39. Hevner, Kate. Experimental studies of the elements of expression in music. *Amer. J. Psychol.*, 1936, **48**, 246–268.
40. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
41. Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
42. Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.
43. Hull, C. L. *Principles of behavior: An introduction to behavior theory.* New York: Appleton-Century, 1943.
44. Jacobson, E. Electrophysiology of mental activities. *Amer. J. Psychol.*, 1932, **44**, 677–694.
45. Johnson, W. Studies in language behavior: I. A program of research. *Psychol. Monogr.*, 1944, **56**, No. 2 (Whole No. 255), 1–15.
46. Jones, H. E. Emotional factors in learning. *J. gen. Psychol.*, 1929, **2**, 263–272.
47. Jones, H. E., & Wechsler, D. Galvanometric technique in studies of association. *Amer. J. Psychol.*, 1928, **40**, 607–612.
48. Jung, C. G. *Studies in word-association.* (Trans. by M. D. Eder.) London: William Heinemann, 1918.
49. Kapustnik, O. P. The interrelation between direct conditioned stimuli and their verbal symbols (trans. from Russian title), 1930. *Psychol. Abstracts*, 1934, **8**, No. 153.
50. Karwoski, T. F., & Berthold, F., Jr. Psychological studies in semantics: II. Reliability of free association tests. *J. soc. Psychol.*, 1945, **22**, 87–102.
51. Karwoski, T. F., Gramlich, F. W., & Arnott, P. Psychological studies in semantics: I. Free association reactions to words, drawings, and

objects. *J. soc. Psychol.*, 1944, **20**, 233–247.

52. KARWOSKI, T. F., & ODBERT, H. S. Color-music. *Psychol. Monogr.*, 1938, **50**, No. 2 (Whole No. 222).
53. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. E. Studies in synesthetic thinking: II. The roles of form in visual responses to music. *J. gen. Psychol.*, 1942, **26**, 199–222.
54. KARWOSKI, T. F., & SCHACHTER, J. Psychological studies in semantics: III. Reaction times for similarity and difference. *J. soc. Psychol.*, 1948, **28**, 103–120.
55. KELLER, MARGARET. Mediated generalization: the generalization of a conditioned galvanic skin response established to a pictured object. *Amer. J. Psychol.*, 1943, **56**, 438–448.
56. KENT, GRACE H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, **67**, 37–96, 317–390.
57. KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt Brace, 1935.
58. KÖHLER, W. *The mentality of apes.* (Trans. by E. Winter.) New York: Harcourt Brace, 1925.
59. KOTLIAREVSKY, L. I. Cardio-vascular conditioned reflexes to direct and to verbal stimuli (trans. from Russian title), 1936. *Psychol. Abstracts*, 1939, **13**, No. 4046.
60. Langfeld, H. S. Note on a case of chromaesthesia. *Psychol. Bull.*, 1914, **11**, 113–114.
61. Lazarus, R. S., & McCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
62. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283–293.
63. LYNCH, C. A. The memory values of certain alleged emotionally toned words. *J. exper. Psychol.*, 1932, **15**, 298–315.
64. McCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205–222.
65. McGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
66. McGINNIES, E. Discussion of Howes' and Solomon's note on "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 235–240.
67. MAIER, N. R. F. Reasoning in humans. III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, **35**, 349–360.
68. MALINOWSKI, B. The problem of learning in primitive languages. Supplement in C. K. Ogden, & I. A. Richards. *The meaning of meaning.* New York: Harcourt Brace, 1938.
69. MARBE, K. *Experimentell-psychologische Untersuchungen über das Urteil.* Leipzig: Engelmann, 1901.
70. MASON, M. Changes in the galvanic skin response accompanying reports of changes in meaning during oral repetition. *J. gen. Psychol.*, 1941, **25**, 353–401.
71. MAUSNER, B., & SIEGEL, A. The effect of variation in "value" on perceptual thresholds. *J. abnorm. soc. Psychol.*, 1950, **45**, 760–763.
72. MAX, L. W. An experimental study of the motor theory of consciousness. III. Action-current responses in deaf-mutes during sleep, sensory stimulation, and dreams. *J. comp. Psychol.*, 1935, 19, 469–486.
73. MAX, L. W. An experimental study of the motor theory of consciousness. IV. Action-current responses in the deaf during awakening,

kinaesthetic imagery, and abstract thinking. *J. comp. Psychol.*, 1937, **24**, 301–344.

74. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific response. *Amer. J. Psychol.*, 1940, **53**, 173–203.
75. Metzner, C. A. The influence of preliminary stimulation upon human eyelid responses during conditioning and during subsequent heteromodal generalization. *Summ. Doct. Diss. Univ. Wis.*, 1942, **7**, 152–154.
76. Miller, G. A. *Language and communication.* New York: McGraw-Hill, 1951.
77. Morris, C. *Foundations of the theory of signs.* Internat. Encycl. Unif. Sci., Vol. 1, No. 2. Chicago: University of Chicago Press, 1938.
78. Morris, C. *Signs, language and behavior.* New York: Prentice-Hall, 1946.
79. Mosier, C. I. A psychometric study of meaning. *J. soc. Psychol.*, 1941, **13**, 123–140.
80. Mowrer, O. H. Preparatory set (expectancy): Some methods of measurement. *Psychol. Monogr.*, 1940, **52**, No. 2 (Whole No. 233).
81. Odbert, H. S., Karwoski, T. F., & Eckerson, A. B. Studies in synesthetic thinking: I. Musical and verbal associations of color and mood. *J. gen. Psychol.*, 1942, **26**, 153–173.
82. Ogden, C. K., & Richards, I. A. *The meaning of meaning.* London: Kegan Paul, 1923.
83. Osgood, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277–301.
84. Osgood, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, **38**, 132–154.
85. Postman, L., & Bruner, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314–324.
86. Postman, L., Bruner, J. S., & McGinnies, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.
87. Razran, G. H. S. Salivating and thinking in different languages. *J. Psychol.*, 1935–1936, **1**, 145–151.
88. Razran, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.
89. Reed, H. B. The existence and function of inner speech in thought processes. *J. exp. Psychol.*, 1916, **1**, 365–392.
90. Riess, B. F. Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26**, 238–240.
91. Riess, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.*, 1946, **36**, 143–152.
92. Ross, R. T. Studies in the psychology of the theater. *Psychol. Rec.*, 1938, **2**, 127–190.
93. Sanford, R. N. The effects of abstinence from food upon imaginal processes: A preliminary experiment. *J. Psychol.*, 1936, **2**, 129–136.
94. Sanford, R. N. The effects of abstinence from food upon imaginal processes: A further experiment. *J. Psychol.*, 1937, **3**, 145–159.
95. Skinner, B. F. The verbal summator and a method for the study of latent speech. *J. Psychol.*, 1936, **2**, 71–107.
96. Skinner, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, **1**, 69–76.
97. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.
98. Stagner, R., & Britton, R. H., Jr. The conditioning technique applied to a public opinion problem. *J. soc. Psychol.*, 1949, **29**, 103–111.

99. STAGNER, R., & OSGOOD, C. E. An experimental analysis of a nationalistic frame of reference. *J. soc. Psychol.*, 1941, **14**, 389–401.

100. STAGNER, R., & OSGOOD, C. E. Impact of war on a nationalistic frame of refefence: I. Changes in general approval and qualitative patterning of certain stereotypes. *J. soc. Psychol.*, 1946, **24**, 187–215.

101. SWITZER, S. C. A. Disinhibition of the conditioned galvanic skin response. *J. gen. Psychol.*, 1933, **9**, 77–100.

102. THORNDIKE, E. L. *The teacher's word book.* New York: Teachers College, Columbia Univ., 1921.

103. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1944.

104. THORSON, A. M. The relation of tongue movements to internal speech. *J. exp. Psychol.*, 1925, **8**, 1–32.

105. THOULESS, R. H. Phenomenal regression to the real object. II. *Brit. J. Psychol.*, 1931, **22**, 1–30.

106. THUMB, A., & MARBE, K. *Experimentelle Untersuchungen über die psychologischen Grundlagen der sprachlichen Analogiebildung.* Leipzig: W. Engelmann, 1901.

107. TRAUGOTT, N. N. The interrelations of immediate and symbolic projections in the process of the formation of conditioned inhibition (trans. from Russian title), 1934. *Psychol. Abstracts*, 1935, **9**, No. 1166.

108. TRAUGOTT, N. N., & FADEYEVA, V. K. The effect of difficult extinction of food-procuring conditioned reflexes upon the general and speech behavior of children (trans. from Russian title), 1934. *Psychol. Abstracts*, 1935, **9**, No. 1167.

109. WARREN, H. C. (Ed.) *Dictionary of psychology.* New York: Houghton Mifflin, 1934.

110. WASHBURN, M. *Movement and mental imagery.* Boston: Houghton Mifflin, 1916.

111. WATSON, J. B. *Behavior, an introduction to comparative psychology.* New York: Holt, 1914.

112. WERTHEIMER, M. *Productive thinking.* New York: Harper, 1945.

113. WITMER, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–360.

114. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

115. WHEELER, R. H., & CUTSFORTH, T. D. Synaesthesia and meaning. *Amer. J. Psychol.*, 1922, **33**, 361–384.

116. WYCZOIKOWSKI, A. Theoretical and experimental studies in the mechanism of speech. *Psychol. Rev.*, 1913, **20**, 448–458.

117. ZIPF, G. K. *The psychobiology of language.* New York: Houghton Mifflin, 1935.

118. ZIPF, G. K. *Human behavior and the principle of least effort.* Cambridge, Mass.: Addison-Wesley, 1949.

Received August 29, 1951.

THE VALIDITY OF THE BLACKY PICTURES[1]

GERALD S. BLUM
University of Michigan
AND
HOWARD F. HUNT
University of Chicago

Just as there is more than one way to skin a cat, there are many ways to try to validate behavior ascribed to a dog. Specific, definitive validity studies in the area of projective techniques are likely to continue to be infrequent by virtue of the fact that the instruments tend to be used in a variety of situations, each use requiring its own demonstration of validity. Conceivably a measuring device may prove quite effective in differentiating clinical syndromes and yet fall down in the assessment of dynamics as they operate within an individual. Or the theory underlying a given technique may make it more applicable to one type of problem than another. Offhand, these and similar complications may account in part for the current phase of confused ardor and recrimination concerning the value of our diagnostic tools.

But perhaps this very diversity may be turned into an asset through a kind of expedient, substitute solution to the criterion problem. While fully adequate criteria continue to be elusive, it seems reasonable to assume that a large variety of partially satisfactory criteria can serve temporarily as the best approximations to the "true" ones. In other words, a technique with merit should show up well when tested in a wide assortment of situations, despite the fact that no one situation provides a sufficient test by itself.

In the past the tendency has been to rely heavily on impressions of clinical utility as a basis for evaluating projective techniques. The Rorschach and Thematic Apperception Tests, for example, seem to have fared very well when judged in this way. While favorable opinions regarding clinical performance certainly deserve consideration, the dangers involved in judging validity solely on such a foundation are self-evident. Here, then, is another source of confusion which contributes to the current tempest over whether the Rorschach and other rojective tests are "valid" or not.

Fortunately, the Blacky Pictures (3) are of such recent vintage that radition and custom do not inevitably interfere with the development of standards for evaluation. Unsystematic, subjective evidence based on clinical impressions need not play much of a role, if any, in the validation process. Of course, it is encouraging to learn that clinicians react positively to the test, just as it would be discouraging if the opposite tended to be the case. But acceptance of a technique as being "fruitful" hardly constitutes validation. An instrument must prove its worth in a variety of research settings.

---

[1] The authors wish to express their appreciation to E. Lowell Kelly for his helpful ggestions in the preparation of this manuscript.

The present article is intended to survey completed studies of the Blacky Pictures in a number of separate areas. These summaries, based on researches reported in detail elsewhere, have been assembled in order to provide some perspective on what has been done with the test to date and to suggest lines of future research. Sufficient evidence has accumulated to warrant such a stock-taking enterprise; and we shall take advantage of the opportunity to present our tentative appraisal of these available validity data.

THE BLACKY PICTURES TECHNIQUE

Before describing the various approaches relating to the validity of the Blacky Pictures, a short résumé of the technique itself is probably in order. The pictures were originally devised in 1947 as an attempt to get at the deeper recesses of personality in a more appropriate setting, geared directly to dynamic interpretation. It was felt that many projective techniques are lacking in theoretical orientation and consequently pose difficult interpretive problems. The twelve cartoons, tailored to fit psychoanalytic theory, portray the adventures of a dog named Blacky. The first introduces the cast of characters, which includes Blacky, Mama, Papa, and Tippy, a sibling figure of unspecified age and sex. Each of the subsequent cartoons is designed to depict either a stage of psychosexual development or a type of object relationship within that development. Dog cartoon figures were utilized to compensate for the increased structuring of the stimulus cards—the intention being to facilitate freedom of personal expression in situations where human figures might provoke an unduly inhibitory resistance as a result of being "too close to home." While on the one hand minimizing the dangers of resistance, the canine medium, thanks to Disney cartoons and comic strips, still preserves enough reality for subjects to identify themselves freely with the cartoon figures and to project their innermost feelings. It seems almost as if the animal cartoons appeal directly to the residues of childish, pre-logical thinking in adults, despite the added fact that they frequently realize consciously that they are telling about themselves.

The subject is asked to tell stories in response to the pictures. After his spontaneous story on each cartoon, he is asked a series of multiple-choice and short-answer questions pertaining to that psychoanalytic dimension. Upon finishing the inquiry on the last card, a third level of response is obtained by having the subject express his preferences for the various cartoons. For group administration the sexes are tested separately, since Blacky is presented as the "son" to males and as the "daughter" to females. An objective scoring system, recently revised on the basis of internal consistency of the four source scores (Spontaneous Story, Inquiry, Cartoon Preference, and Related Comments) for each dimension, is employed for research uses of the instrument. Subjects wind up with a profile of scores in terms of three possible degrees (0, +, ++) of disturbance on every dimension. When dichotomization of scores is desirable, the + and ++ categories are combined.

Comparison With Theoretical Predictions

The fact that the Blacky Pictures are based on psychoanalytic theory opens one kind of proving-ground for testing the instrument. The theory yields predictions which can be compared directly with results from the test. Completed studies of this type fall into three categories: syntactical relationships within the theory itself; application of the theory to demographic groups; and theoretical contributions to the understanding of various clinical syndromes. Research findings pertaining to the first two categories have already been published in detail (2) and will be summarized again only briefly. Four illustrations from the last category will then be described below.

Syntax of the Psychoanalytic Theory of Psychosexual Development

In the initial investigation, the Blacky Pictures were administered, in group form, to 119 male and 90 female students in elementary psychology classes at Stanford University. Analysis of the resulting protocols revealed statistically significant correlations (beyond chance expectancy) between dimensions on the test, which suggests that the dimensions are related to each other in a nonrandom manner. These significant correlations agreed with expectations based upon the internal syntax of the theory in those instances in which a definite prediction could be deduced from the theory.

Application of the Theory to Demographic Groups

As part of the same study, the Blacky protocols were also analyzed for sex differences. The number of obtained differences again clearly exceeded the total that would be expected to occur by chance alone.

These differences between the sexes were almost all in the direction predicted by psychoanalytic theory in those instances for which the theory permitted an unequivocal prediction.

The Theory as Applied to Various Clinical Syndromes

The paranoid schizophrenic. Aronson (1), in a study aimed at the Freudian theory of paranoia, administered the Blacky Pictures individually along with several other psychological tests to 30 paranoid schizophrenics (Pa), 30 nonparanoid schizophrenic controls (Ps), and 30 normal controls (N). The three groups were equated for age, intelligence, occupation, religion, and veteran status. The two psychotic samples were differentiated according to the presence or absence of delusions, which was judged on the basis of all the available clinical materials. We shall summarize here only the section dealing with the Blacky Pictures. Significance tests computed from the blindly-scored, coded protocols revealed a large number of differences (including source scores and inquiry items as well as total dimensional scores) between the groups. The number expected to occur by chance alone could not be computed, since the various comparisons were not independent of each other (e.g., total dimensional score contained the separate sources). However, since the Blacky data were compared to predictions made from the theory, the issue of capitalization on chance was not deemed to be of major concern.

Within each test dimension (Oral Eroticism, Oral Sadism, etc.) attempts were made to integrate the findings, taking into account the level of awareness tapped by separate sources (Spontaneous Story, Inquiry, etc.). The following are excerpts from Aronson's summaries of the data (1, pp. 129–130):

> 1. *Oral Eroticism* (Prediction: Pa stronger than N). The paranoids showed more evidence of basic underlying oral deprivations than the normals. However, they tended, much more than the normals, to deny these deprivations. These results are in accord with the Freudian contention that paranoids are fixated at an early narcissistic stage of psychosexual development.
>
> 2. *Oral Sadism* (Prediction: Pa stronger than N or Ps). The paranoids exhibited very strong unconscious oral sadistic conflicts, as compared to the normals, but they did not differ from the psychotics. The paranoids also showed a great deal of difficulty in handling oral sadistic impulses at a near conscious level. These results indirectly support Freud's statement that paranoids are fixated at a narcissistic stage of development. They directly support Klein and Rosenfeld's theory that oral sadistic fixations play an extremely important role in the development of paranoia.
>
> 3. *Anal Retentiveness* (Prediction: Pa stronger than N or Ps). The paranoids revealed much stronger anal retentive conflicts than the normals did. They also tended to exceed the psychotics in this respect. These results are in agreement with the prevalent analytic viewpoint that anal fixations are the primary ones in paranoia. The paranoids made little attempt to conceal their anal retentive

interests and attitudes. This frankness on their part, in contrast to their general evasiveness on the Blacky technique, was probably due to the fact that anal retentive character traits are rather highly valued in Western society, and hence do not have to be denied. Finally, according to analytic theory, anal fixations frequently lead to the adoption of passive (homo)sexual aims.

Summarizing the remaining dimensions, he found the paranoid group showing greater evidences of masturbation guilt; conscious attempts at denial of strong underlying castration anxiety; a consistent tendency toward feminine identification; severe superego conflicts; and a preference for narcissistic types of love-object choice. Little or no differences were obtained on the dimensions of Anal Expulsiveness, Sibling Rivalry, and Ego Ideal. Some indirect evidence of stronger oedipal fixations in the paranoid group appeared on the Anaclitic Love-Object dimension, but not on Oedipal Intensity.

The stutterer. Merchant (8) utilized the Blacky Pictures to explore the psychosexual development of stutterers. Blacky protocols (group administration) were obtained from 20 stutterers and 20 non-stutterers attending a middle-western college. Every stutterer was matched with a non-stuttering student of the same sex, age, family constellation, and classification on the Ohio College Aptitude Test. There were 15 men and 5 women in each group. She found that the stutterers differed significantly from the controls on 7 of the 14 Blacky dimensions. The experimental group showed more evidence of disturbance on Oral Eroticism, Oral Sadism, Castration Anxiety and Penis Envy, Identification Process, Guilt Feelings, and Anaclitic Love-Object. On the dimension of Anal Expulsiveness the stutterers received significantly fewer strong scores. Merchant evaluates her findings as follows:

> The Blacky Test presents a fairly clear-cut picture of the psychosexual development of these stutterers, and one in agreement with psychoanalytic theory of stuttering according to Fenichel. Developmental status of stutterers is sufficiently consistent, or the Blacky Pictures sufficiently sensitive, or both, that these characteristics did show up with only 20 pairs of subjects. . . .
>
> The final paragraph of such a pilot study must be one of "Suggestions" rather than of "Conclusions," but it may suggest rather strongly that here is evidence for the psychoanalytic theory of stuttering and for the Blacky Pictures as an instrument useful in the investigation of psychosexual development.

The peptic ulcer case. In a study reported by Blum and Kaufman (4) the Blacky Pictures were administered, in group form, to 14 adult male ulcer patients as one phase of a total psychological workup. The results are summarized as follows:

> Exploration of the scored responses, in conjunction with those of three control groups (normals, paranoid schizophrenics, and nonparanoid schizophrenics), uncovered two opposite trends within the ulcer sample. Whereas all the ulcer cases wrote strongly oral stories, only one-half selected multiple-choice

alternatives indicating oral conflict in the inquiry items on Cartoon I. The discrepancy between close-to-conscious expression of oral needs in the one half, contrasted with obvious attempts at denial in the other, suggested the hypothesis that there may be two very different patterns of ulcer dynamics. Further investigation of the "primary" and "reactive" subgroups showed the latter to be consistently more evasive in the content of their stories on the first oral cartoon. Comparisons on the remaining Blacky dimensions revealed the "primary" pattern to consist of a relative absence of conflicts centering about anal retentiveness, oedipal feelings, castration anxiety, and narcissistic object-choice, along with an excessive concern for mothering and little conscious desire for strong, masculine identification. The "reactive" pattern, closer to the popular conception, is characterized by repressed receptive longings, intense, unresolved oedipal feelings, the conscious wish to emulate a decisive father figure, a pervasive sense of guilt, and a highly narcissistic approach to others.

The formulation of these two patterns, based upon a limited number of cases, was presented in the form of a suggestion for the direction of additional research in the area of peptic ulcer dynamics. Differences between the two subgroups on other psychological tests are reported in a paper by Marquis, Sinnett, and Winter (**7**).

The sexual offender. Lindner[2] reports what he describes as a "validating study" of the Blacky Pictures done in two Maryland State prisons. The technique was administered individually to 67 male sexual offenders and 67 nonsexual offenders matched according to age, race, I.Q., education, marital status, socioeconomic status, length of sentence, and previous commitments. The experimental group was comprised, in decreasing order of frequency, of cases of homosexuality, sodomy, rape, pedophilia, exhibitionism, carnal knowledge, and contributing to the delinquency of minors. Analysis of the data revealed the sexual offenders to be significantly more disturbed than the nonsexual offenders on 9 of the 13 Blacky dimensions. These groups were differentiated at the .001 level of significance on Oral Eroticism, Oedipal Intensity, Masturbation Guilt, Castration Anxiety, Sibling Rivalry, and Guilt Feelings; at the .01 level on Oral Sadism and Narcissistic Love-Object; and at the .05 level on Anaclitic Love-Object. Anal Expulsiveness also approached significance ($P < .10$), whereas Anal Retentiveness, Positive Identification, and Positive Ego Ideal did not discriminate between the sexually deviant group and the nonsexually deviant controls. Within the experimental sample of offenders there were no significant differences according to classification.

Lindner presents the following interpretation of his findings:

> In conclusion, then, we have found that the Blacky Pictures test is a valid indicator of psychosexual deviation in a selected population. It is sufficiently

---

[2] LINDNER, H. A study of imprisoned sexual and nonsexual offenders by means of the Blacky Pictures test. (Presented at the 1951 annual meeting of the Eastern Psychological Association.)

sensitive to discriminate between two groups of subjects: a sexually aberrant group and a nonsexually aberrant group. To the extent that this test represents psychoanalytic theory, these data may be considered to support such theory as a plausible rationale.

VALIDATION BY EXPERIMENTAL TECHNIQUES

The preceding section summarized a series of studies in which Blacky data were compared with predictions derived from psychoanalytic theory. A second type of approach to the validity of the instrument lies in comparisons with independent experimental measures. This method is illustrated by the case of a nine-year-old girl who served as a subject in a broad research program aimed at the psychoanalytic theory of the "oral character" (5, 6). As one phase of the research, the Blacky Pictures were administered to all 26 children in the fourth grade of an elementary school in Ann Arbor, Michigan. In the process of checking the adequacy of test administration a Blacky record (all identifying marks removed) was randomly selected for a supervisory consultation. The stories told by this girl revealed very strong involvement on Cartoon I (Oral Eroticism) and spontaneous oral references on seven of the remaining ten pictures. The striking evidence of preoccupation with orality suggested a handy way to validate her test performance, since data for the class were available on a large number of experimental measures of oral character structure.[3] These measures had been obtained from experiments, time-sampling, teacher ratings, and sociometrics.

Out of a total of 20 variables our heroine fell in the most oral third of the group on 16 variables. On 12 of the 20 variables she ranked between first and fifth. For example, she was fifth in the class of 26 on average number of one-ounce cups of ice cream consumed daily; first on teacher ratings of "needs praise from others"; third in taste suggestibility; third in the inability to say "No" to requests from others; and so on. It is also interesting to note that on two nonoral control variables—neatness and collecting tops of cups during the ice cream experiment—she ranked 23rd and 13th respectively.

The fact that independent experimental evidence clearly substantiated the blind Blacky prediction of strong oral involvement in this girl should encourage such approaches to validation.

---

[3] BLUM, G. S. Validation of a child's Blacky record by experimental techniques. (Presented at the round table on the Blacky Pictures held at the 1950 annual meeting of the American Psychological Association.)

PREDICTION OF BEHAVIOR IN A GROUP SETTING

Swanson (**10**), in an attempt to analyze the relationship between personality characteristics of group members and their observed behavior in the group, administered the Blacky Pictures to each of two groups of 20 delegates at the 1949 session of the National Training Laboratory for Group Development in Bethel, Maine. These groups contained roughly equal numbers of men and women, with a median age of 47. Independent data on the behavior of participants and groups were secured as part of the general routine of the project and were not obtained specifically for this study.

The Blacky protocols were later scored in a blind analysis by the author of the test, who had available only the age and sex of the subject. A profile of scores along the various psychoanalytic dimensions was thus available for each subject. The three conventional categories of very strong, fairly strong, and weak or absent were employed, so that one individual might have gotten a ++ on Oral Eroticism, a + on Oral Sadism, a 0 on Anal Expulsiveness, and so on down the line. Training Groups A and B were both scored in this fashion.

Swanson, from his knowledge of psychoanalytic theory, made predictions concerning the pattern of Blacky dimensional scores which should be associated with various tendencies. For example, he reasoned that strong scores on Oral Sadism, Anal Expulsiveness, Oedipal Intensity, Sibling Rivalry, and Guilt Feelings should raise the total amount of an individual's actual participation in a permissive group like A, whereas Oral Eroticism, Anal Retentiveness, and Anaclitic Love-Object choice should lower it. Using this intuitive pattern analysis of Blacky scores, he ranked the members of the group on predicted participation and then correlated the rankings with observed actual participation. The patterns of predictions based on Blacky scores for Group B had to be revised somewhat in view of its markedly different atmosphere. In contrast to Group A, which was extremely permissive and productive in its decision-making, B had a climate of hostility, which was accompanied by lower participation and less effective decision-making.

The main results of the study are presented in Table 1, which is reproduced from Swanson's report. In commenting on the utility of the Blacky Pictures for his design, he writes:

> Unsystematic though our data are, they suggest that the Blacky Test is a suitable instrument for recording the kind of personality variables that are important in interaction. Neither the number of our groups, nor their size, nor the

variety of predictions attempted, permit anything like a final verdict on its adequacy, but it seems to merit further use when such an instrument is needed in research designs.

What has been especially impressive in working with these personality variables is the utility of an interpersonal interpretation of psychoanalytic theory. Along with its technical ingenuity, much of the merit of the Blacky Test seems to rest on the fact that, unlike most other personality depth instruments, with such exceptions as the Rorschach, it originated in a systematic theory that tries to handle the individual mechanisms of interpersonal relations in a single, comprehensive scheme. Unquestionably that scheme is not infallible, but it is systematic and testable.

TABLE 1

SUMMARY OF RELATIONSHIPS BETWEEN THE BLACKY PREDICTIVE PATTERNS AND THE CRITERION VARIABLES FOR GROUPS A AND B*

After Swanson (10)

Criterion Variable	*Clinical Prediction Based on Blacky Scores*	
	Group A	*Group B*
Total volume of participation	.63§	.56†
Percentage of total participation expressed that is disagreement and modification	.21	.32
Percentage of total participation received that is disagreement and modification	.33	.57‡
Percentage of total liking choices that are favorable	.68§	.49†
Percentage of total mentions of being influenced that are positive in direction	.36	.75§
Percentage of total mentions of being a source of influence that are positive in direction	.52‡	.55‡
Ideology Test—total score	.56‡	.33
Desire to be a leading member of the group	.68§	.37
Desire to conform to the group's wishes	.50‡	no data
Satisfaction with own performance in the group	†‡	no data

* Except in the one case noted, all measures in this table are rhos. N is 17 in each group.

† Relationship measured by the use of simultaneous probability.

‡ Significant at the .05 level.

§ Significant at the .01 level.

THE CLINICIAN'S JUDGMENT AS A STANDARD OF COMPARISON

The clinical approach, in which the clinician himself fulfills the role of criterion agent, is a more conventional procedure in general use. To the authors' knowledge, no systematic Blacky explorations of this sort have already been completed. However, we may be able to set our sights more sharply by considering three "exercises" which fall into this category.

The therapist's judgment as the standard. One such investigation has been discussed in the preface to the sample case workup in the Blacky Pictures Manual.[4] The procedure, in which a patient's therapist was asked after 10 hours of therapy to pick out his patient from among a group of five, is described as follows:

> All of the patients had been tested and the analyses written up blindly before therapy was begun. The therapist made choices first from five short descriptions (laid out in random order) adapted from each dimension, and finally from the five complete personality evaluations. On all of the specific dimensions he succeeded in identifying his patient by either first or second choice, and in the final analysis he made the correct first choice (**3**, p. 8).

A similar study along more extensive lines is now being conducted by Ellis in which the Blacky Pictures have been administered to a patient with some two hundred hours of psychoanalysis.[5] Ratings by therapist and patient, without knowledge of Blacky results, are being compared with ratings made by 30 clinical psychologists from the Blacky protocol alone.

The clinician's prediction of the patient's test performance. A variation in the use of the clinician as a criterion agent is to ask him to predict, on the basis of all available information, how the patient will perform on the test. An illustrative exercise was done at VAH Fort Custer in which an extensive set of interview and test materials (Rorschach, TAT, Bender-Gestalt, Wechsler-Bellevue) was used to predict presence (+) or absence (0) of disturbance on each of the Blacky dimensions.[6] The predicted and actual scores for this case coincided in 12 out of 13 instances insofar as gross indication of presence or absence of disturbance was concerned.

The clinical case study. Michal-Smith, Hammer, and Spitz (**9**) report the case study of a nine-year-old Negro boy referred to a clinic for stealing toys. Case material revealed a history of spying on his mother and stepfather at night, masturbation, homosexual and heterosexual episodes. Excerpts from the patient's Blacky stories dramatically confirm and elaborate the severe oedipal ambivalence expected on the basis of the history.

---

[4] The clinician whose cooperation made this exercise possible was Dr. Abraham Carp, Chief Psychologist, VAH, Fort Custer, Michigan.

[5] Personal communication.

[6] The clinician in this case was Justin L. Weiss, VA trainee in clinical psychology, University of Michigan.

SOME CRITICAL COMMENTS

At first glance, the research cited above might be interpreted as pretty clearly establishing the validity of the Blacky Pictures. For various reasons, however, these results should be considered as encouraging and suggestive rather than definitive. They encourage further exploration of both the test and the personality theory it serves; they strongly suggest that "there is something there," but do not necessarily indicate "what it is," or "where it is." Two examples from the original monograph illustrate methodological difficulties which constrain a definitive interpretation:

> 1. The only theoretical predictions which could be compared with empirical findings were those predictions which, on the basis of their content, appeared to be related to the differences or correlations obtained on the test. How many predictions could the theory generate? And what proportion of this hypothetical population of predictions was actually checked and found to agree with data from the test? Nobody can tell; psychoanalysis is still insufficiently logical and systematic to permit an exhaustive determination of all possible, meaningful predictions. The theory was carefully combed and those predictions worked out which seemed pertinent to the dimensions of the test. Some, of course, might have been missed. In addition, only statistically significant differences and correlations in the data from the test were compared with predictions from the theory; nonsignificant differences and correlations, though they might have been quite legitimately reflecting absence of relationship or difference, had to be left out of the comparisons because of their ambiguity (2, p. 23).
>
> Thus we see that only certain kinds of data from the test entered the comparisons with what was probably only an incomplete population of theoretical predictions. The total possible number of comparisons between theory and test is unknown, so a rigorous determination of what constitutes chance agreement between theory and test cannot be made. In these circumstances, one cannot determine, precisely, whether obtained agreement between theory and test significantly exceeds the "real" chance expectancy for agreement, though there is no question that it exceeds the most reasonable estimate of expectancy that we can make at the moment.
>
> 2. For another example, let us turn to the test. How should one interpret the obtained statistically significant sex differences and dimensional intercorrelations? Is each significant difference "real" because it is statistically significant? Are all of them "real" just because more than a chance number of them seemed to occur? Not necessarily. The statistical analysis in this first study could do no more than indicate an apparent excess of such significant differences; it could not indicate which were "real" and which attributable to chance. For this reason, all must be considered to have only a tentative sort of status until cross-validating studies, on other samples, reveal which differences and correlations survive the vicissitudes of sampling and random error, and which do not. In effect, the results must be interpreted as hypotheses to be tested in later studies under a variety of conditions.

One may find the seeds of ad hoc reasoning, capitalization on chance, and kindred sins in the subsequent researches summarized above. These methodological frailties frequently, often unavoidably, characterize exploratory research in this area—hence the emphasis here on the suggestive and encouraging, but not definitive, character of the current findings on the validity of the Blacky Pictures. It is noteworthy, however, that these studies investigated the test performance of various kinds of persons, ranging from psychiatric cases to normal individuals, and employed criteria as diverse as psychiatric symptomatology, behavior in a group, and amount of ice cream eaten. In a very important sense, these studies constitute a species of cross-validation of the test; and, fortunately, they all tend to yield positive and confirmatory results. As such evidence accumulates, one's confidence in the test can justifiably increase sharply. It becomes a bit difficult to consider that sampling errors, capitalization on chance, and the like, by themselves, are responsible for the positive results obtained under such varied circumstances.

Of course, the full range of possibilities for validation research has by no means been explored as yet. Some studies should employ, more directly, the type of data upon which psychoanalytic theory is based—the events of analytic therapy. Up to the present, no systematic investigation has compared performance on the pictures with such criteria as predictions and judgments made by trained analysts, the course and content of the patient's associations, the vicissitudes of the transference and so on. Since these are the data psychoanalysis was invented to order and comprehend, test and theory could then be compared at a somewhat less derivative and inferential level than has been possible heretofore. Many of the relevant and even critical aspects of personality may be detected less in the form of overt behavior on the part of the patient than through hunches, feelings, impressions, and even personal reactions on the part of the therapist. More refined research in this area demands the use of therapists and observers as sensitive discriminators. With appropriate study, these sources could yield criterion data relevant to those inchoate, unexpressed aspects of personality so dimly and ambiguously reflected in overt symptoms, social behavior, and the like.

The close articulation of the Blacky Pictures with psychoanalytic theory may or may not prove to be a great virtue. Both are in their scientific infancy; both really are still in the "invention" stage of development. If the theory, or something like it, turns out to be essentially correct, the advantages of having a device like the Blacky are obvious. If the theory proves to be grossly wrong in either metaphor or syntax, Blacky's prognosis also will be bleak, unless he succeeds in demonstrating his discriminatory power in experimental, behavioral studies of personality which do not depend entirely upon psychoanalytic theory.

At present, one cannot be sure of the ultimate fate of psychoanalysis as a coherent, logical, and systematic theory. Nor can one forecast Blacky's future with certainty. The instrument needs much additional research and exploration. Its early publication for general use was designed to facilitate such activity and to broaden its base. In the present article we have sought to convey recent findings and to achieve some perspective in their evaluation.

REFERENCES

1. Aronson, M. L. A study of the Freudian theory of paranoia by means of a group of psychological tests. Unpublished doctoral dissertation, University of Michigan, 1950.
2. Blum, G. S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, **3**, 3–99.
3. Blum, G. S. *The Blacky Pictures: A technique for the exploration of personality dynamics.* New York: Psychological Corp., 1950.
4. Blum, G. S., & Kaufman, Jewel B. Two patterns of personality dynamics in male ulcer patients, as suggested by responses to the Blacky Pictures. *J. clin Psychol.*, in press.
5. Blum, G. S., & Miller, D. R. Exploring the psychoanalytic theory of the "oral character." Presented at the 1950 APA symposium entitled "Experimental Approaches to Psychoanalytic Theory." *J. Pers.*, in press.
6. Blum, G. S., & Miller, D. R. Exploring the psychoanalytic theory of the "oral character": a follow-up study. *Amer. Psychologist*, 1951, **6**, 339–340. (Abstract)
7. Marquis, D. P., Sinnett, E. R., & Winter, W. D. A psychological study of peptic ulcer patients. *J. clin. Psychol.*, in press.
8. Merchant, Frances C. Psychosexual development in stutterers. *J. consult. Psychol.*, in press.
9. Michal-Smith, H., Hammer, E., & Spitz, H. Use of the Blacky Pictures with a child whose oedipal desires are close to consciousness. *J. clin. Psychol.*, 1951, **7**, 280–282.
10. Swanson, G. E. Some effects of member object-relationships on small groups. *Human Relations*, in press.

Received August 25, 1951

A MEASURE OF RELATION DETERMINED BY BOTH MEAN DIFFERENCE AND PROFILE INFORMATION

CHARLES E. OSGOOD AND GEORGE J. SUCI
University of Illinois

In psychology and other social sciences the investigator often has available a number of different measurements, constituting a profile, for each of several variables. The relationship between any two such variables is usually expressed by a product-moment correlation or some derivative of this ratio. This expression of relationship reflects only the *profile* similarity between the two sets of measures; information regarding the absolute *differences*, or discrepancies, between the means of the measurements is discarded. In many cases this difference information is irrelevant—when, for example, one correlates height (measured in inches) with weight (measured in pounds), or when one correlates two aptitude tests in which the scores depend quite arbitrarily upon the number of items in each battery. But there are many other cases where this difference information is relevant. Measurement specialists have often expressed concern, for example, over the fact that a moron and a genius may have closely correlated profiles, despite the absolute discrepancy between their scores. During the past three years the writers have been working toward the development of objective methods of measuring connotative meaning, using systems of descriptive scales as the bases for judgments.[1] As will be shown in subsequent examples, usual correlation procedures failed to give valid representation of semantic relations, and we were forced to search for other measures of relationship. The *difference method* we have developed, while particularly applicable to semantic measurements, is by no means restricted to this type of data. Problems identical with ours exist in many areas of social science,[2] and probably elsewhere as well.

---

[1] OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, 49, 197–237. A preliminary description of the difference method reported here was given in two papers at the meeting of the American Psychological Association, September, 1950.

[2] As evidence of concern over such problems we may cite the following: R. B. Cattell (r_p and other coefficients of pattern similarity. *Psychometrica*, 1949, 14, 279–298) discusses exactly the same issue in connection with the measurement of differences in culture patterns, deriving a "coefficient of pattern similarity" which is similar to the method described here; Cronbach has recently described (mimeographed materials) a measure of relation which appears to be identical with that described here, but developed in connection with interpersonal measurements from sociometric and Q-technique studies.

THE DIFFERENCE METHOD AND THE GEOMETRIC MODEL ON WHICH IT IS BASED

Before describing the difference method it will be useful to clarify the nature of the problem. Figure 1 shows five seven-point scales. In our own work on semantic measurement such scales are defined by polar adjectives. When subjects are instructed to rate each of a number of concepts on each of the scales (e.g., LADY *rough–smooth*, SIN *high–low*, FATHER *strong–weak*, etc.), profiles like those shown are generated. Concepts are represented by capital letters (A, B, C, etc.) and scales for judgment by lower-case letters (a, b, c, etc.). Suppose A represents LOVE, B represents AFFECTION and C represents HATE. Since they covary perfectly, despite the gross absolute discrepancies, intercorrelations among them would all be 1.00, leading to the inference that LOVE is as similar in meaning to HATE as it is to AFFECTION. The correlation of any other concepts with D would be indeterminate since the variance of D is zero, yet concept D (e.g., AGGRESSION) is clearly closer to C (e.g., HATE) than to either A or B. Thus it can be seen that the product-moment correlation may not only distort the information but may be completely inapplicable in some cases. What is needed for these and similar kinds of data is a measure of relationship which takes into account both the mean discrepancies and the profile

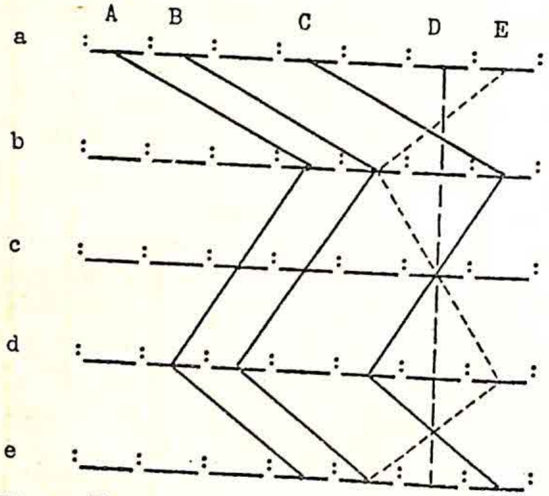


FIG. 1. PROFILES FOR FIVE HYPOTHETICAL VARIABLES (A, B, C, ETC.) JUDGED ON FIVE SCALES (a, b, c, ETC.).

VARIABLES A, B, AND C CONSTRUCTED SO AS TO HAVE IDENTICAL PROFILES; VARIABLES C, D, AND E CONSTRUCTED SO AS TO HAVE NEARLY IDENTICAL MEAN DIFFERENCES FROM A.

covariation, thereby providing fuller reflection of the total information in the measurements.

If we have n variables and k measurements, every variable can be considered a point fixed in k-dimensional space by the k coordinates. Data of this kind can be arranged in matrix form as shown in Figure 2.

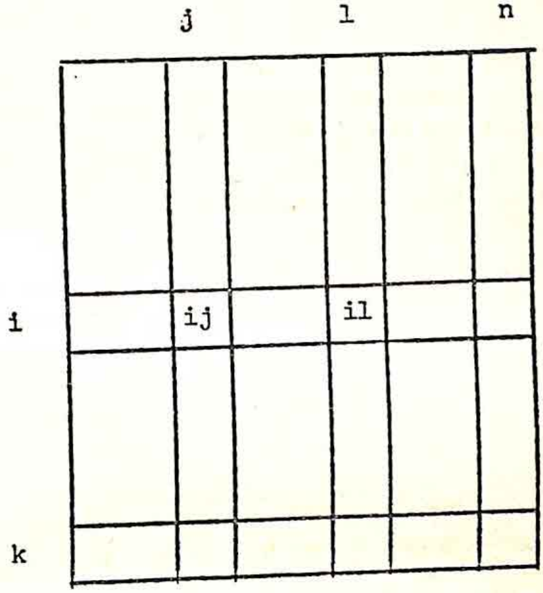


FIG. 2. GENERALIZED SCORE MATRIX FOR n VARIABLES MEASURED ON k DIMENSIONS.

Let us define the following symbols:

X = a measurement in cell ij in raw score form;
$\underset{j}{z}$ = a measurement in cell ij standardized with respect to all measures in column j;
$\underset{i}{z}$ = a measurement in cell ij standardized with respect to all measures in row i;
d_{jl} = the difference between the score in cell ij and the score in cell il;
D = the distance between variables j and l;
$D(z)$ = this distance D found with z scores in the matrix;
$D(X)$ = this distance D found with raw scores in the matrix.

If the k dimensions are mutually orthogonal, the distance D between any two points fixed in this space is given by the square root of the sum of the squared differences between coordinates on the same dimension, i.e.,

$$D = \sqrt{\sum_i d_{jl}^2} \qquad [1]$$

This distance, D, directly provides a measure of the relationship between variables j and l. For n variables one may then find $n(n-1)/2$ values of D. Since D has a direct geometrical counterpart,[3] the n variables can be plotted in space such that the distances between each variable and every other variable satisfy the values in the matrix of D. Actual physical representation of such relations is limited to cases where $k=3$, but the values in the D matrix are accurate for n dimensions (however, see discussion below).

In Table 1 are given X and $D(X)$ values derived from the hypothetical profiles drawn in Figure 1. The values for D indicate that variable A is less closely related to C than it is to B: ($D(X)$ for A, C equals $\sqrt{45}$ while $D(X)$ for A, B equals $\sqrt{5}$); the correlational measures are 1.00 in both cases. Profiles C, D, and E are drawn so as to have equal mean discrepancies with respect to A, yet so as to have radically different patterns of covariation with A. *Under these conditions, D takes into account profile relations.* This measure of relationship is minimal (e.g., the variables are closest together) when there is parallel covariation (for A, C, $D(X)$ equals $\sqrt{45}$); the distance increases for unrelated profiles (for A, D, $D(X)$ equals $\sqrt{58}$); the distance is maximal, and the indexed relationship least, when profiles have perfect negative correlation (for A, E, $D(X)$ equals $\sqrt{72}$). Thus D, as a measure of relationship, takes into account both the absolute discrepancy between sets of measurements as well as their profile similarities. The operations

TABLE 1

RAW DATA (X) AND DISTANCE MEASURES ($D(X)$) FOR A SAMPLE OF VARIABLES (A, B, C, ETC.) LOCATED ON A SAMPLE OF SCALES

Scale	X				
	A	B	C	D	E
a	1	2	4	6	7
b	4	5	7	6	5
c	3	4	6	6	6
d	2	3	5	6	7
e	4	5	7	6	5

	$D(X)$				
	A	B	C	D	E
A					
B	$\sqrt{5}$				
C	$\sqrt{45}$	$\sqrt{20}$			
D	$\sqrt{58}$	$\sqrt{31}$	$\sqrt{7}$		
E	$\sqrt{72}$	$\sqrt{45}$	$\sqrt{21}$	$\sqrt{4}$	

---

[3] In our original work we had thought this method somehow dependent on correlational matrices. It was only after discovering that the method applied to raw score data that a mathematical rationale became apparent. We are grateful to Professor Jozef Cohen for pointing out the relationship between our D and the geometrical model.

for finding D have the additional advantage of being extremely simple and non-time-consuming. One simply sums the squared differences, one row at a time, for each pair of variables and looks up the square root of each such sum in the usual tables. Using calculating equipment, this is essentially a single operation upon the matrix.

Relations to Factor Analysis

The existence of a geometrical model for the D matrix immediately suggests a possible relation to the factor analytic model. Suppose that the variables j and l are part of a "battery" of variables whose intercorrelations have been factor analysed, and suppose further that this analysis indicates two factors. The factor loadings are coordinates on the two factors which fix the two points in two-dimensional space. These two points represent the termini of two vectors whose lengths are h_j and h_l respectively. The relation between j and l can be expressed independently of the coordinates by the correlation r_{jl}, which in the factor analytic model is the scalar product $h_j h_l \cos \phi_{jl}$, where h is the square root of the communality and ϕ_{jl} is the angle between the vectors j and l. By the law of cosines, the distance, V (to distinguish it from D), between the termini of the two vectors j and l is found from the relation:

$$V = \sqrt{h_j^2 + h_l^2 - 2h_j h_l \cos \phi_{jl}} = \sqrt{h_j^2 + h_l^2 - 2r_{jl}}. \qquad [2]$$

If we now transform the raw score matrix (X) to standardized scores $\underset{j}{z}$ and compute the $D\underset{j}{z}$ matrix, we can write the following relations:

$$D(z) = \sqrt{\sum_i \left(\underset{j}{z} - \underset{l}{z}\right)^2} = \sqrt{\sum_i \underset{j}{z}^2 + \sum_i \underset{l}{z}^2 - 2\sum_i \underset{jl}{zz}}\,. \qquad [3]$$

Since,

$$\frac{\sum_i \underset{j}{z}^2}{k} = \frac{\sum_i \underset{l}{z}^2}{k} = 1.00, \quad \text{and} \quad \frac{\sum_i \underset{jl}{zz}}{k} = r_{jl}, \qquad [4], [5]$$

$$D(z) = \sqrt{k + k - 2kr_{jl}} = \sqrt{2k(1 - r_{jl})}. \qquad [6]$$

When

$$h_j^2 = h_l^2 = 1.00,$$

$$V = \sqrt{2(1 - r_{jl})}$$

therefore,

$$D(z) = \sqrt{k}\, V. \tag{8}$$

Thus, under certain conditions—standardization with respect to all measures for the variables and all communalities equal to unity—a linear relation exists between distances in the D model and distances (V) in the factor analytic model. It should be noted that the operation of standardizing measurements eliminates information regarding mean differences exactly as do the operations followed in computing product-moment correlations.

TABLE 2

DATA OBTAINED FROM ONE SUBJECT JUDGING 10 CONCEPTS ON A SET OF 20 SEMANTIC SCALES

Semantic Scale	A	B	C	D	E	F	G	H	I	J	
	QUICKSAND	WHITE ROSE BUDS	DEATH	HERO	METHODOLOGY	FATE	VIRILITY	GENTLENESS	SUCCESS	SLEEP	
high	6	3	7	1	3	6	2	4	1	5	low
green	4	2	4	7	5	4	7	1	6	1	red
weak	7	1	6	7	5	5	7	1	7	3	strong
rough	5	7	2	3	2	2	3	7	5	7	smooth
active	7	7	7	1	3	7	1	7	1	7	passive
empty	7	5	1	7	5	1	7	7	7	6	full
small	7	1	5	7	2	4	7	4	7	5	large
cold	1	3	1	7	5	2	7	5	7	3	hot
clear	7	4	7	1	5	6	1	6	1	7	hazy
young	5	1	7	1	1	7	1	4	2	4	old
good	7	1	7	1	1	5	1	2	1	3	bad
peaceful	4	1	5	5	3	5	5	1	4	1	ferocious
sick	3	7	1	7	5	3	7	4	7	6	healthy
angular	7	7	3	6	3	5	6	7	5	7	rounded
tense	6	7	6	5	2	5	2	7	4	7	relaxed
sad	1	6	1	7	5	2	6	5	7	6	happy
soft	3	1	4	5	3	2	7	1	6	1	loud
wet	2	4	5	5	3	5	3	4	3	3	dry
beautiful	7	1	7	2	3	5	2	1	2	2	ugly
fresh	6	1	7	1	1	6	1	3	1	2	stale

Application to Sample Data

One subject rated 10 arbitrarily chosen concepts (such as HERO, FATE, GENTLENESS, and SLEEP) against the set of 20 seven-point scales shown in Table 2. $D(X)$ values by the difference method as well as intercorrelations between all concepts were computed. The matrix of intercorrelations was factor analysed, yielding three factors. The raw score data are given in Table 2. The matrix for $D(X)$, obtained by the operation already described, is shown as Table 3. Table 4 gives the factor matrix and Table 5 the means and variances for the same

TABLE 3

Distance Measures ($D(X)$) Relating the 10 Concepts Given in Table 2

	A	B	C	D	E	F	G	H	I	J
A										
B	16.70									
C	9.32	19.10								
D	17.75	14.90	20.27							
E	15.30	11.45	16.13	10.25						
F	9.45	15.33	5.29	17.52	13.12					
G	17.70	16.73	20.40	4.36	9.90	17.83				
H	13.60	6.86	16.61	15.40	12.33	13.68	16.43			
I	17.40	14.56	20.22	3.74	9.95	17.64	3.87	14.80		
J	12.10	7.21	15.72	15.30	12.21	12.61	16.22	4.58	14.56	

TABLE 4

Factor Loadings and Communalities for the Variables Given in Table 2

Variable	*Factor Loading*			*Communality*
	I	II	III	h^2
A	−.44	.37	.42	.51
B	.59	.73	−.07	.93
C	−.93	.16	.29	.97
D	.81	−.43	.25	.90
E	.62	−.19	.16	.45
F	−.79	.31	.24	.78
G	.77	−.53	.26	.94
H	.37	.78	−.09	.83
I	.85	−.43	.21	.95
J	.35	.82	.16	.82

TABLE 5

MEAN SCALE POSITIONS AND VARIANCES FOR THE VARIABLES GIVEN IN TABLE 2

Variable	*Mean*	σ^2
A	5.10	4.3
B	3.50	6.1
C	4.65	5.4
D	4.30	6.4
E	3.25	2.1
F	4.35	3.1
G	4.15	6.6
H	4.05	5.1
I	4.20	5.9
J	4.30	4.8

original data as in Table 2. Certain relationships are evident from inspection of Tables 3 ($D(X)$ values) and 4 (factor loadings): variables having similar loadings on the three factors are close together in terms of the D distances (compare variables D, G and I).

Figure 3 is a drawing of the three-dimensional model constructed

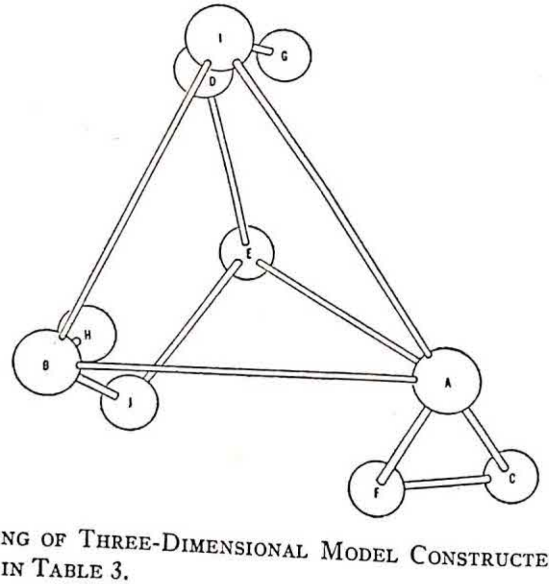


FIG. 3. DRAWING OF THREE-DIMENSIONAL MODEL CONSTRUCTED FROM DISTANCE MEASURES GIVEN IN TABLE 3.

VARIABLES D, G, I (HERO, VIRILITY, SUCCESS), B, H, J (WHITE ROSE BUDS, GENTLENESS, SLEEP), A, C, F (QUICKSAND, DEATH, FATE), AND E (METHODOLOGY).

from the $D(X)$ values in Table 3. In constructing these models each $D(X)$ value, as a distance on some convenient linear scale, is conceived as the radius of a sphere: variable B may be placed anywhere on a sphere 16.70 units from variable A; variable C must fall somewhere on the circle defined by the intersection of two spheres, 9.32 units from A and 19.10 units from B; variable D must fall at one of the two points of intersection of three spheres with radii of 17.75, 14.90, 20.27 from A, B, and C respectively. The position of variable F is completely determinate within three dimensions, as are the positions of all remaining variables. In practice we use small rubber balls to represent the variables and thin "pick-up" sticks to maintain correct distances. The model in Figure 3 specifies three clearly isolated clusters of variables: A, C, F (QUICKSAND, DEATH, and FATE), B, H, J (WHITE ROSE BUDS, GENTLENESS, and SLEEP), and D, G, I (HERO, VIRILITY, and SUCCESS). The remaining variable, E (METHODOLOGY) is independent of these clusters. The validity of this empirical categorization of these concepts is obvious.

Figure 4 shows the relation betwen $D(X)$ distances and V, distances between the end points of vectors derived from the factor matrix for the same data, as a correlational scattergram. The close relation is apparent. The encircled points, falling somewhat away from the linear relation, all involve either variable A or variable E. Inspection of Table 4 indicates that these variables are most deviant from 1.00 in their values for h^2; inspection of Table 5 indicates that these variables are also among those most deviant in terms of means and variances. This is an empirical demonstration of the relation between $D(X)$ and V derived above, in this case employing unstandardized raw data. Although we have applied the difference method to relations among concepts in the present instance, the same procedure can be used to determine relations among the scales of judgment (tests) or among the subjects who do the judging.

SOME PROBLEMS CURRENTLY BEING INVESTIGATED

One problem appears in the illustration used above: Why, when 20 scales are used for judgments, do the concepts plot in three dimensions? It is certain that these 20 scales are not all independent;[4] they are oblique to varying degrees and may fall within a three-dimensional space. With oblique axes the distance formula proposed above probably distorts the "true" relation, but the nature and amount of such distortion is debatable. If x and y are two orthogonal axes and a set of points plotted on them falls in a straight line, then x and y are perfectly related, the straight line has reduced the dimensionality from two to one, and the relations among the points are not distorted. When conditions are not extreme—when dimensions are neither independent nor perfectly correlated—there may be distortion of an as yet undetermined magnitude. Distortion will definitely occur if one factor is represented by a large number of the k dimensions whereas the remaining dimensions are mutually independent. For the present, and for the purposes of our semantic measurements, we are solving these problems by finding a set of scales known to be independent from the results of factor analysis.

---

[4] Factor analysis of 50 descriptive scales, from which these 20 were drawn, indicates that many of them are closely interrelated.

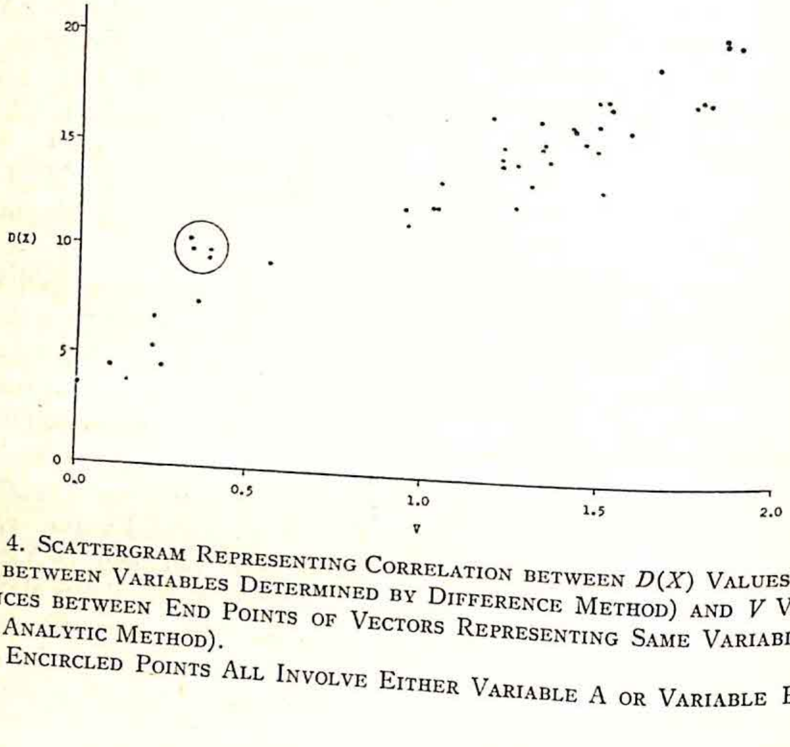


Fig. 4. Scattergram Representing Correlation between $D(X)$ Values (Distances between Variables Determined by Difference Method) and V Values (Distances between End Points of Vectors Representing Same Variables in Factor Analytic Method).

The Encircled Points All Involve Either Variable A or Variable E (see Text).

Another problem concerns the units of the k dimensions, which must be comparable if D distances are to be interpretable. In the case of our semantic measurements this is automatically taken care of by the format of the scales and instructions to the subjects. With other kinds of data, e.g., personality tests scores, the units may be different from measurement to measurement. The value D would be affected by this inconstancy and the results liable to misinterpretation. In such cases it is possible to standardize or rank scores across variables for each measurement. The score matrix would then consist of $\underset{i}{z}$ and the D matrix of $D\underset{i}{z}$. The judgment of the investigator will determine when such transformations are necessary.

We have not as yet developed adequate methods for determining the signficance of differences between distances. Although the clusters displayed in Figure 3 are self-evident, this is not always the case and we need a method for determining when a variable does or does not fall within a cluster. Similarly, in comparing the structures of variables obtained from an individual or group at different times (e.g., when an experimental factor such as psychotherapy has been introduced), we should like to be able to estimate the significance of changes in relationship. Chi-square and F seem the most promising statistics of those now available, but it may prove necessary to develop tests directly from the logic of this method itself.

Finally, there are two phenomena for which we have no adequate rationale, but which have empirical validity. First, simple summation of absolute differences, $\sum_i |d_{jl}|/k$, results in a matrix of distances which plots just as accurately as $\sqrt{\sum_i d_{jl}^2}$. Furthermore, these distances relate to factor analytic results just as well as do those in a $D(X)$ matrix. Whether these results are artifacts of the kinds of numbers we have used or are due to some mathematical relation has not yet been determined. Second, when D is found from a matrix of intercorrelations, the results yield very accurate representations of the factor analytic pattern for the same variables. Using the full symmetrical matrix of correlations and entering the main diagonal with the highest correlations in each column (Thurstone's estimate of communality), D's are obtained by exactly the same operation as described above. When V's (from factor loadings) and D's (from correlation matrix) were computed for Thurstone's box and cylinder problems,[5] the correlations between these two sets of distance estimates were .97 and .94 respectively.

SUMMARY

A method for analyzing interrelationships among variables has been described which takes into account not only the profile similarity among the sets of measures but also their mean differences. This measure, D, can be applied to matrices of raw scores and, apparently, to correlational matrices as well. It has the advantage of an existing geometrical model which permits graphic representation of the data in many cases. It has the further advantage of unusual simplicity and speed in procedure. It has been shown to yield results identical to those of factor analysis under certain specified conditions. But the chief value of this method is not as an approximation to factor analysis, but as an independent method which does not eliminate information regarding differences between the means of the related variables. A number of unsolved problems are currently being studied.

---

[5] THURSTONE, L. L. *Multiple-factor analysis*. Chicago: University of Chicago Press, 1947.

Received for priority publication March 5, 1952.

ERRATUM

An error occurred in the citing of various references in the article "Thematic Apperception Test: Interpretive Assumptions and Related Empirical Evidence," by Gardner Lindzey, in the January 1952 issue of THIS JOURNAL. The correction is as follows:

On page 1, the reference "Lawrence (**39**)" should read "Douglas (**18**)." With the exception of this reference, throughout the article author references numbered from 18 through 38 should have one number added to them to give the correct item in the reference list, e.g. Freud (**24**) should read Freud (**25**).

A NOTE ON FUNCTIONAL RELATIONS OBTAINED FROM GROUP DATA

MURRAY SIDMAN
Columbia University

The empirical determination of functional relations between behavior and its controlling variables forms a large part of modern behavioral research. One important aspect of this type of experimentation is the method of distributing subjects among the various points which determine an empirical curve.

The most direct method is to use a single organism, and the same organism, to obtain every point on the curve. This procedure is not always practicable, however, for one or both of two reasons.

1. Intra-organism variability may be so great as to obscure any lawful relation. It is sometimes possible to avoid this problem by taking several determinations at each point and using a statistical measure, a common technique in obtaining threshold measurements (2).
2. Even this procedure will not be effective if, as is often the case, the experimental operations involved in determining one point on the curve have an effect upon the values of other points. For example, one cannot use the same organism to determine all the points on a function relating extinction responding to number of reinforcements. One reason for this is that the extinction operation is itself a variable entering into extinction results subsequent to reconditioning (4). It is seldom, if ever, possible to get around this difficulty by using a different individual for each point on the curve. Here inter-organism variability comes into the picture to obscure lawfulness.

Faced with these problems, most experimenters turn to group data. One technique is to employ the same group of organisms to obtain all the points. This procedure, however, is also ruled out if the second situation mentioned above is in effect (unless, of course, this is the problem under investigation). The only recourse remaining is to use a different group to determine each point. The rest of this paper will be devoted to a discussion of certain considerations involved in this latter method of obtaining an empirical function.

INDIVIDUAL VS. AVERAGED FUNCTIONS

The first point to be made is that the mean curve obtained by such a procedure is not necessarily of the same shape as the inferred individual curves. (The term "inferred" is used here since this method is generally employed when the individual curves cannot be obtained directly.) The following development brings this out clearly. For the purpose of demonstration we take as our example the negatively accelerated positive growth function which has achieved a certain prominence in behavior theory. This function can be expressed as

$$y = M - Me^{-kx}, \tag{1}$$

where M is the asymptote approached by y, and k determines the rate of approach to M. If the curves for individual organisms are of this shape, inter-organism variability might occur in the asymptotes approached by the curves, in the rates of approach to the asymptotes, or both. Figure 1 represents a set of individual curves which vary with respect to both constants. (Although, for the sake of simplicity in Figure 1, M and k are assumed to be positively correlated, this assumption is not necessary.) When, for the reasons mentioned above, it is not possible to obtain these individual curves empirically, the procedure generally followed is to expose a different sample of the population of subjects to each value of the independent variable, x, and to take the mean of the dependent variable as the corresponding value of y. On the assumption that each of the samples is equally representative of the population, this procedure is represented in Fig. 1 by the broken

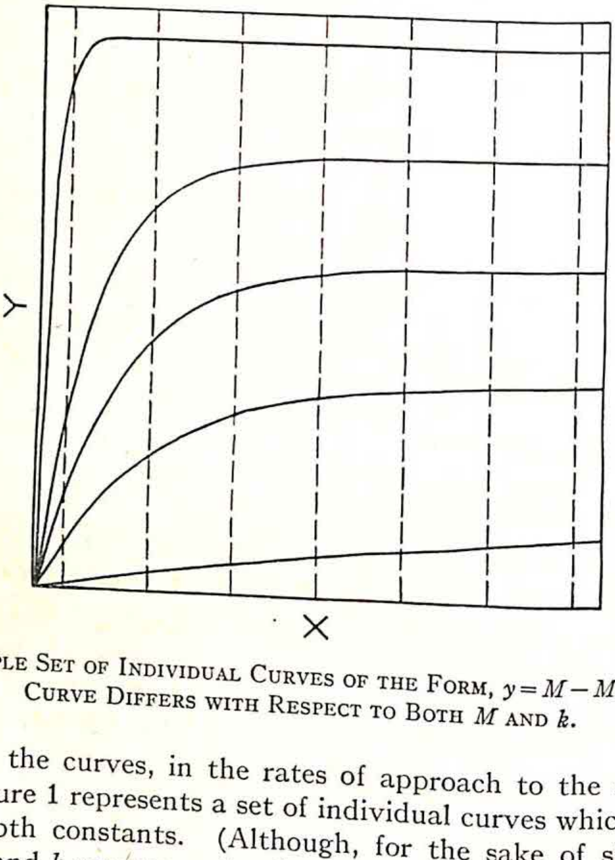


FIG. 1. SAMPLE SET OF INDIVIDUAL CURVES OF THE FORM, $y = M - Me^{-kx}$. EACH CURVE DIFFERS WITH RESPECT TO BOTH M AND k.

lines drawn from selected values of x. These lines simply indicate that, in a given experiment, the distribution of functions is cut through at selected values of the independent variable.

Corresponding to the curves of Figure 1, we can write the following equations:

$$\begin{aligned} y_1 &= M_1 - M_1 e^{-k_1 x} \\ y_2 &+ M_2 - M_2 e^{-k_2 x} \\ &\cdot \\ &\cdot \\ &\cdot \\ y_n &= M_n - M_n e^{-k_n x}. \end{aligned} \tag{2}$$

To determine the mean value of y for a given value of x we sum equations [2] and divide by n, which results in the expression,

$$n^{-1} \sum_{i=1}^{n} y_i = \bar{y} = n^{-1} \left(\sum_{i=1}^{n} M_i - \sum_{i=1}^{n} M_i e^{-k_i x} \right), \tag{3}$$

which can be written

$$\bar{y} = n^{-1} \left[\sum_{i=1}^{n} M_i - (M_1 e^{-k_1 x} + M_2 e^{-k_2 x} + \cdots + M_n e^{-k_n x}) \right]. \tag{4}$$

Each of the exponentials in

$$S = M_1 e^{-k_1 x} + M_2 e^{-k_2 x} + \cdots + M_n e^{-k_n x} \tag{5}$$

can be expanded to give

$$\begin{aligned} S = {} & [M_1 + M_1(-k_1)x + M_1(-k_1)^2 x^2/2! + M_1(-k_1)^3 x^3/3! + \cdots] \\ & + [M_2 + M_2(-k_2)x + M_2(-k_2)^2 x^2/2! + M_2(-k_2)^3 x^3/3! + \cdots] \\ & + \cdots \\ \cdots & + [M_n + M_n(-k_n)x + M_n(-k_n)^2 x^2/2! + M_n(-k_n)^3 x^3/3! + \cdots]. \end{aligned} \tag{6}$$

Upon rearranging coefficients we have

$$\begin{aligned} S = {} & \sum_{i=1}^{n} M_i + [M_1(-k_1) + M_2(-k_2) + \cdots + M_n(-k_n)]x \\ & + [M_1(-k_1)^2 + M_2(-k_2)^2 + \cdots + M_n(-k_n)^2]x^2/2! + \cdots \\ & + [M_1(-k_1)^m + M_2(-k_2)^m + \cdots + M_n(-k_n)^m]x^m/m! + \cdots. \end{aligned} \tag{7}$$

This can be expressed

$$S = \sum_{i=1}^{n} M_i + A_1 x + A_2 x^2/2! + \cdots + A_m x^m/m! + \cdots, \tag{8}$$

where

$$A_m = M_1(-k_1)^m + M_2(-k_2)^m + \cdots + M_n(-k_n)^m. \qquad [9]$$

Substituting equation [8] into equation [4], we arrive at

$$\bar{y} = [1-(1 + A_1x + A_2x^2/2! + \cdots + A_mx^m/m! + \cdots)]n^{-1}. \qquad [10]$$

Equation [10] will reduce to the form [1] if and only if

$$A_i = A_j \qquad [11]$$

for all i, j, or if

$$A_i = A_1{}^i. \qquad [12]$$

Condition [11] is impossible, since the A's are alternately negative and positive. Condition [12] is easily demonstrated to be impossible unless the A's all equal unity, in which event [12] becomes a special case of [11].

It has been shown, then, that for individual curves of the form [1], if inter-organism variability occurs both in the asymptotes and in the rates of approach to these asymptotes, the average curve *cannot* be described by an equation of the form [1].[1] It can be seen from equations [7] to [10] that this will also be the case if the asymptotes are equal and variability occurs only in the rates of approach. Only when the rates of approach are equal will the mean curve be of the form [1]. Thus, under the assumption that

$$k_i = k_j \qquad [13]$$

for all i, j, equation [4] can be rewritten

$$\bar{y} = n^{-1}\left(\sum_{i=1}^{n} M_i - \sum_{i=1}^{n} M_ie^{-kx} \right). \qquad [14]$$

As far as this writer is aware, the assumption that the rates of approach to the asymptotes are equal for all the organisms in a given experiment has never been explicitly acknowledged by any experimenter or theorist who has fitted this growth function to data obtained by the method under discussion.[2]

At this point it may be argued that although the mean curve is not the same as the individual curves, it is similar enough that, within the limits of experimental error, it can be fitted satisfactorily by the same function. Although this argument possesses dubious merit on grounds of theoretical consistency, it can also be attacked by demonstrating that many other types of individual curve will, if averaged, give as good an approximation to [1] as will equation [10].

---

[1] The author is indebted to Mr. L. A. Gardner, Jr. for the essential elements of this demonstration.

[2] Hull appears actually to have made the opposite assumption. He states, "The 'constant' numerical values appearing in equations representing primary molar behavioral laws vary . . . from individual to individual . . . " (**3**, Postulate 18).

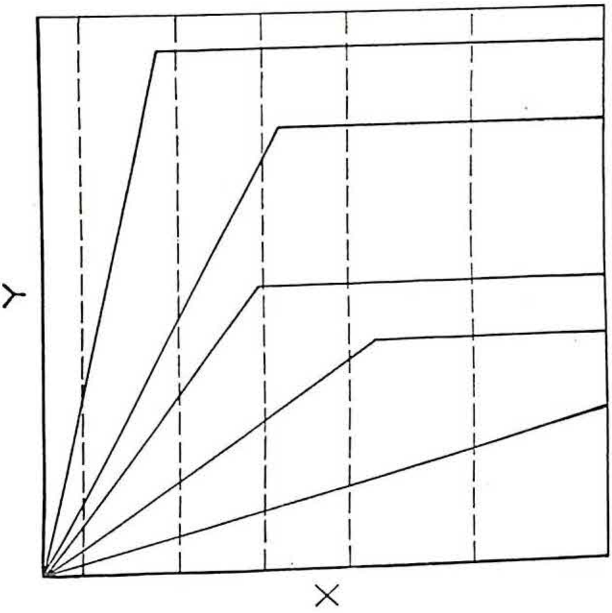


FIG. 2. SAMPLE SET OF INDIVIDUAL CURVES OF THE FORM $y = mx$ UP TO A POINT OF DISCONTINUITY, AFTER WHICH $y = c$.

For example, if the individual curves are straight lines of the form

$$y = mx \tag{15}$$

up to a given value of x, at which point there occurs a discontinuity (see Fig. 2) after which

$$y = c, \tag{16}$$

it can be shown that the mean curve can be described by

$$\bar{y} = f(x)(1 - ax) + aBx, \tag{17}$$

where B is the sum of the maximum values of y and a is a proportionality constant between the slope, m, and the maximum value, y_c, of y. (This assumption of proportionality is not necessary, but is made merely to simplify the discussion.) $f(x)$ is a function describing the relationship between x and the sum of the y_c which have been reached at any value of x. (It can be seen from Fig. 2 that, as x increases, more of the individual curves will reach their maximum values of y.) $f(x)$ will be determined by the frequency distribution of y_c and the relation, if any, between y_c and x. The form of $f(x)$ will determine how closely equation [16] approximates equation [1]. We see, then, that if the individual curves are of the form indicated in Figure 2, the mean curve may approximate equation [1] or any one of a large number of other forms.

Although this discussion has treated only two specific cases in any detail, the same type of analysis can be carried out for any functional relation. In some cases it will be found that the mean curve will be of the same form as the individual curves, e.g., the straight line. However, many functions will show the property discussed above, namely, that the mean curve cannot be of the same form as the individual curves except under special conditions. Furthermore, given a particular mean curve, the form of the individual curves is not uniquely specified. It appears, then, that when different groups of subjects are used to obtain the points determining a functional relation, the mean curve does not provide the information necessary to make statements concerning the function for the individual.

ALTERNATIVE PROCEDURES

Given a situation like that outlined above, there are several alternatives open to experimenters and theorists. First the suggestion might be made that all data obtained by the averaging procedure outlined above be ignored and that the questions which such data attempt to answer not be asked. This radical solution is probably not necessary. Such mean curves may give some valuable information, depending upon the validity of the assumptions one is willing to make concerning the general lawfulness of individual behavior. If it is assumed that all individuals of a certain class will display the same type of functional relation in a given situation, then the mean curve will tell something about that function. If, under this assumption, we obtain the mean curve described by equation [10], it will be known that the individual curves are increasing functions of x and that they either reach a maximum or approach an asymptote.

However, we are not forced to make such an assumption. The mean curve of equation [10], for instance, can be obtained even if the individual curves are so irregular that they cannot be described by any useful equation. A more profitable approach might be to obtain and present all data in the form of distributions, to specify the distributions by their form and by their parameters, and to relate these distributions to the independent variables. (Such procedures would, of course, apply also to experiments in which the same group of subjects is used to determine all points on a function, but where it is observed that the individual data are not amenable to a functional description.)

A third alternative is to develop techniques which will produce lawful individual functions and to present the data without averaging. Although many methods have already been developed for work with individuals (e.g. **1**, **5**, **6**, **7**), many more would have to be worked out either by devising new measuring techniques or by attaining more rigorous control over extraneous variables. Statistical procedures would enter into these methods in at least two ways. Replicative statistics would be necessary to determine the reliability of the individual curves, and information would probably be needed concerning the population distribution of the curve constants. Even if these methods were highly developed, however, there are still some data (such as speed of acquisition of behavior under different motivating conditions) which will probably never be amenable to individual treatment. It will, in such cases, be necessary to forego such data or to use statistical analysis. If the latter is done there remains the problem of theoretical integration of data obtained by two different procedures. A decision among the alternate choices will be made only on the basis of further empirical investigation.

REFERENCES

1. Frick, F. C., & Miller, G. A. A statistical description of operant conditioning. *Amer. J. Psychol.*, 1951, **64**, 20–36.
2. Graham, C. H., & Veniar, F. A. The influence of size of test field surround on visual intensity discrimination. *Proc. nat. Acad. Sci.*, 1950, **36**, 17–25.
3. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
4. Perkins, C. C., Jr., & Cacioppo, A. J. The effect of intermittent reinforcement on the change in extinction rate following successive reconditionings. *J. exp. Psychol.*, 1950, **40**, 794–801.
5. Skinner, B. F. *The behavior of organisms: an experimental analysis.* New York: Appleton-Century, 1938.
6. Skinner, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193–216.
7. Skinner, B. F., & Campbell, S. L. An automatic shocking-grid apparatus for continuous use. *J. comp. physiol. Psychol.*, 1947, **40**, 305–307.

Received August 20, 1951.

A NOTE ON THE COMBINATION OF RATINGS ON THE BASIS OF RELIABILITY

C. H. LAWSHE AND BRYANT F. NAGLE
Occupational Research Center, Purdue University

In psychological research one is often faced with the question of how to combine ratings when a measure of rater reliability is the only attribute of the ratings known. This situation arises frequently in setting up a criterion when there is no way in which to determine the validity of the various ratings. The question becomes one of how to combine the ratings of the different judges on the basis of their reliability.

In partial answer to this problem, Shen (**9**) developed a formula for estimating the reliability of an individual rater from the intercorrelation matrix of the raters. Kelley (**2**) has followed up Shen's work by producing a table of weights that may be used when combining the ratings. With these tools, psychologists have been able to go in at least four directions: (a) weight the raters equally; (b) eliminate a rater with low reliability and weight the remaining ones equally; (c) weight the raters according to the weights given by Kelley; and (d) eliminate a rater with low reliability and weight the remaining ones according to their Kelley weights. This paper represents an empirical look at the results these four approaches will give.

The data used in comparing the methods come from actual studies. The five cases used are briefly described:

Case 1. Six industrial executives rated 46 general foremen in a tractor plant on the basis of how well the men were doing their jobs. The type of rating used was paired comparison, utilizing partial pairings as suggested by McCormick and Bachus (**4**) and McCormick and Roberts (**5**). From Nagle (**8**).

Case 2. Five supervisors rated 14 office jobs on the basis of which should carry the highest pay rate. Paired comparison ratings were used. From Miles (**7**).

Case 3. Four professors rated 23 graduate students in industrial psychology on their ability to do graduate work. Paired comparison was used. From Lawshe and Deutsch (**3**).

Case 4. Four supervisors rated 24 journeyman proofreaders on their versatility in proofreading. Paired comparison was used. From McGinley (**6**).

Case 5. Three graduate students in psychology rated 40 responses to a picture on a positive-negative attitude expressed towards the theme represented by the picture. A five-point rating scale was used. From Boyce (**1**).

Data drawn from these five cases are presented in Table 1. In columns 2 and 3 the number of raters and the number of rated objects are shown. Shen's formula was applied to obtain a reliability estimate for each rater, and the estimate for the least reliable judge is presented in column 4. Column 5 shows the mean Shen reliability[1] of all judges in the case.

As indicated above, one approach involves the computation of the mean of all the possible intercorrelations between the judges and the subsequent stepping up of this mean by the Spearman-Brown formula.

TABLE 1

Estimated Reliability of Ratings with Various Numbers of Judges Using Equal and Unequal Weightings

Case	*No. Raters*	*No. Things Rated*	*Lowest Shen Reliability*	*Average Shen Reliability*	*Ave. Intercorrel. of Judges*		*Spearman-Brown Reliability*		*Weighted Composite Reliability*	
					N Judges	$N-1$ Judges	N Judges	$N-1$ Judges	N Judges	$N-1$ Judges
1	2	3	4	5	6	7	8	9	10	11
1	6	46	.304	.57	.549	.605	.880	.884	.890	.887
2	5	14	.568	.79	.800	.860	.952	.961	.955	.972
3	4	23	.394	.58	.585	.643	.849	.844	.860	.852
4	4	24	.695	.91	.825	.862	.950	.949	.986	.951
5	3	40	.189	.44	.400	.543	.667	.704	.701	*

* Shen reliabilities cannot be computed where there are only two raters.

The resulting statistic yields an estimate of the pooled ratings of all judges with individual ratings equally weighted. Column 6 shows the average of all of the intercorrelations for each case and column 8 shows the corresponding stepped up values.

If, on the other hand, we do not use all of the judges but in each case drop out the single judge with the lowest Shen estimate (the one whose reliability is reported in column 4) and compute the mean intercorrelation of the remainder, we obtain slightly higher values. These values are reported in column 7 in the table. When these ratings are pooled (weighted equally) the reliability of the composite may be estimated by the Spearman-Brown values as shown in column 9.

Kelley's recommended procedure (weighting ratings *unequally* in relation to reliability) yields composite ratings, the reliability of which may be estimated by means of the following formula:[2]

---

[1] Computed after conversion to Fisher's z.

[2] For this formula the authors are indebted to Mr. James Norton, Division of Educational Reference, Purdue University. It was derived from the formulas used by Kelley to determine weights for reliability coefficients.

$$r_w = \frac{\sum r_i W_i^2 + 2\sum r_{ij} W_i W_j}{\sum W_i^2 + 2\sum r_{ij} W_i W_j}$$

where r_i = Shen reliability of any given rater i,
W_i = Kelley weight for rater i,
W_j = Kelley weight of any rater except i,
r_{ij} = correlation between rater i and rater j.

If all raters have equal Shen reliabilities, this formula will give the same composite reliability estimate as the Spearman-Brown formula.[3] If raters have differing reliabilities and consequently different weights this formula will always yield a higher estimate than the Spearman-Brown. It should be emphasized that the two estimates cannot legitimately be compared statistically because they are based on different assumptions. Column 10 contains the Shen-Kelley estimates using all raters, whereas entries in column 11 were obtained by eliminating the rater with the lowest reliability (i.e., the one listed in column 4), recomputing rater reliabilities and weights, and using the above formula.

The comparisons that should be made from this table are column 8 with 9, 10 with 11, 8 with 10, and 9 with 11. The first two comparisons (8 with 9, and 10 with 11) show the effect of eliminating a rater whose reliability is considerably below the average of the others. It will be noted that in the first four cases the differences between the reliabilities of N and N-1 (columns 8 and 9) are slight, and there is no consistent tendency for the N-1 reliabilities to be higher or lower than those of N. Case 5 shows a larger difference. No technique is known for estimating the statistical significance of this difference, but from a practical significance point of view it must be admitted that the difference is small.

A comparison between columns 8 and 10, and between 9 and 11 suffers from the fact that the computation of the entries in columns 8 and 9 employ different assumptions from those in 10 and 11. But some sort of comparison, "with reservations," can be made based on the fact that the Spearman-Brown procedure and the weighted composite reliability formula do yield the same results when raters are weighted equally and that when not equally weighted, the latter procedure results in a larger estimate. Inspection of the coefficients shows that the differences

[3] Since Fisher's z transformation should not be used in computing Shen reliabilities this is true only when Fisher's z transformation is *not* used in computing the average of the matrix of rater intercorrelations. The authors have employed Fisher's z in obtaining the values in columns 6, 7, 8, and 9. The differences between the values in columns 8 and 9 and those which would have been obtained had Fisher's z transformation not been used range from .000 to .013, the median being .003.

between the reliabilities are very small, none exceeding .036, though all the weighted composite reliabilities (columns 10 and 11) are larger as expected.

Within the limitations of the comparison made here, it can be seen that the decision to eliminate a low-reliability rater from a rater pool or to apply differential weights to different ratings is not too fruitful. The authors are forced to conclude that much of the time and effort spent in determining the reliabilities of raters and combining their ratings using differential weights is for naught. Even where there are wide differences in the reliabilities of the raters, the elimination of the very low rater does not seem to improve the composite reliability. In fact, the composite reliability may on occasion be lowered! When weights based on the reliabilities are applied to the ratings, no practical improvement in composite reliability appears to be effected.

Because we do not wish to be accused of taking a "pot shot" at some of our more statistically sophisticated colleagues, one further comment seems desirable. The statistician's job is one of finding new ways of making evaluations, of seeking more sensitive tests of significance, and of finding more nearly optimum ways of weighting and combining data. The applied psychologist must, however, reserve the prerogative of deciding how much refinement is enough. There is probably no substitute for empirical evidence in deciding whether or not a particular procedure is more precise than the nature of the data can justify.

REFERENCES

1. Boyce, J. E. Reliability of ratings of picture responses. Unpublished master's thesis, Purdue Univer., 1951.
2. Kelley, T. L. *Fundamentals of statistics.* Cambridge: Harvard Univer. Press, 1947, 419–425.
3. Lawshe, C. H., & Deutsch, S. The interests of industrial psychology students. *J. appl. Psychol.*, 1952, **36**, in press.
4. McCormick, E. J., & Bacchus, J. A. Paired comparison ratings: 1. The effect on ratings of reduction in the number of pairs. *J. appl. Psychol.* (accepted).
5. McCormick, E. J., & Roberts, W. K. Paired comparison ratings: 2. The reliability of ratings based on partial pairings. *J. appl. Psychol.*, in press.
6. McGinley, A. D. The investigation of criteria of job performance in a printing plant. Unpublished master's thesis, Purdue Univer., 1950.
7. Miles, Minnie C. Studies in job evaluation: 9. Validity of a check list for office jobs. *J. appl. Psychol.*, 1952, **36**, 97–103.
8. Nagle, B. F. A study of foreman responsibility. Unpublished master's thesis, Purdue Univer., 1951.
9. Shen, E. The reliability coefficient of personal ratings. *J. educ. Psychol.*, 1925, **16**, 232–236.

Received August 18, 1951.

Tolman, Edward Chace. *Collected papers in psychology.* University of California Press, 1951. Pp. xiv+269. $4.50.

The bibliography at the end of this volume lists 82 titles, dating from 1917 to 1949. From this list an anonymous group of colleagues and former students have selected nineteen representative items for republication. The volume begins with "A New Formula for Behaviorism" (1922) and ends with "Cognitive Maps in Rats and Men" (1948). It is prefaced by a brief but excellently written appreciation of Tolman as scientist and as teacher. Tolman himself had nothing to do either with its initiation or with the selection. His only direct contribution to the book, apart from its content, is a footnote to the Foreword, in which he characteristically expresses the wish "that a mere half of the kind things said had been true." On this point the editors need no reassurance. They might have said many more kind words without violating the traditions of academic restraint. They deserve our gratitude, not merely for having paid fitting tribute to one of the great men of contemporary psychology, but also for having presented a volume which in itself is an important contribution to the history of recent psychological thought.

For more than thirty years Tolman has been in the thick of the battle, a battle in which it is sometimes difficult to tell who is fighting against whom, but in which the participants, especially Tolman, thrust and slash at each other with every evidence of enjoyment. The editors were probably wise in weighting the selection with polemical items, for, although Tolman has contributed generously to our store of psychological facts, he is at his liveliest when he marshals his forces—his rats, his neologisms and his amazing diagrams—and charges into the debate. For a convinced pacifist he has been involved in an extraordinary number of fights. It is characteristic, however, that he always apologizes before he knocks his opponent down and apologizes again as he helps him to his feet.

This book might well serve as "core" reading for a course in Systematic Psychology, for it gives us the history of a system coming into being and grappling in turn with virtually all the fundamental problems with which a psychological system must deal. Whether or not one agrees with Tolman, it is a refreshing and enlightening experience to look at the problems of psychology through Tolman's eyes. The reviewer will perhaps be pardoned if, instead of reviving the old controversies, he selects at random a few of the impressions with which the book has left him.

1. Tolman is not ashamed of, or afraid of, our philosophical heritage. For

him the persistent problems of psychology are the problems of cognition and of motivation. In the last analysis these are the problems of epistemology and of ethics. Tolman has battled stoutly for what he considers to be an objective and scientific psychology, but he has always considered his particular science as merely one of many approaches to an understanding of the totality of things. Tolman speaks to, and listens to, the philosophers. The rest of us might be better psychologists if we followed his example.

2. Much of Tolman's systematic writing has been in the form of protest; but behind every protest is an affirmation. He protested against the "mentalists," against the Titchenerian introspectionists, against teleology. He took issue with Watson and the muscle-twitchers, with Kuo and the anti-hereditarians, and with those who saw in the conditioned reflex a magic formula that would solve all psychological problems. His constant affirmation has been that psychology can and must be an "objective" science. The criterion of objectivity favors behavior rather than experience as the primary subject-matter of psychology, but it does not lead him to deny or even to disparage immediate experience. In fact, he asserts, as a phenomenologist would, that immediate experience "contains quite as much objectivity as it does subjectivity" and, as Titchener would, that it is "an initial common matrix out of which both physics and psychology are evolved" (p. 96). His behaviorism reflects the belief that, if psychology is to be a science, its facts and laws must be stated in a language that is not uniquely related to the experience of any one individual. Where Tolman differs from many other behaviorists is in his stubborn insistence that this language be rich enough to encompass every significant variable in behavior (and experience). He wants his behaviorism to deal with "real life," not just with *S-R* artifacts. Purpose, for example, is a fact of observation; therefore we must have words like means-end-readiness and sign-gestalt-expectation. This reviewer thinks he has a fairly good reading knowledge of the language, but he confesses that he is sure of its meaning only when he has found a counterpart for each word in his own experience.

3. Tolman is not one to retreat before a new idea, or an old one. He advances to meet it, with a gleam in his eye and another diagram to put on the board. In his early writings he was worried about the problems set by the psychologists of an older generation, by James, Titchener, McDougall, and Watson. Then he went to Vienna, and one has the feeling that there he really enjoyed himself. On the one hand, there were the Vienna positivists, who were thinking operationally, and there was Egon Brunswik, whom Tolman later welcomed into his laboratory in Berkeley. On the other hand, there was the lingering tradition of Brentano and act psychology. In all his subsequent writings one can recognize both types of operational emphasis—on the operational criterion in scientific definition, on the operation (act) as the subject-matter of psychology. Tolman's system has grown to meet the challenge of the Gestaltists and of Kurt Lewin. As the Freudian psychology gained in popularity he did not by-pass it like so many of his colleagues. He accepted the challenge and tried to find a place in operational behaviorism for the dynamics of the unconscious. (The diagrams at this point become quite bewildering.) If the book had continued beyond 1948, the reader would have had a chance to observe Tolman grappling with the fundamental concepts of sociology.

4. Probably no psychologist has ever learned so much from rats. Tolman loves them, he is really curious about them, and he believes that the essential variables of human behavior are to be found in the rat. He intends, as he says on p. 163, "to go ahead imagining how, *if I were a rat*, I would behave . . . ." But does the rat really furnish an adequate paradigm for the solution of the human riddle? Being a modest man, Tolman might admit that human behavior presents a greater variety of psychological phenomena; but he would probably argue that a sound training in rat psychology enables one to see the human problems more clearly. Perhaps he is right. To this reader it seems clear, however, that Tolman could never have been such a good rat psychologist had he not been a good human psychologist to begin with. The myopic, anosmic behaviorist of the Watsonian era looked at the rat and saw nothing but muscle twitches. Tolman looked at the rat and saw much more, and animal psychology has been the richer for it.

One minor criticism, and a very friendly one. A volume of "Collected Papers" faintly suggests that a career has come to an end. How about a subtitle: "Tolman in Midpassage"?

Cornell University. ROBERT B. MACLEOD.

KATONA, GEORGE. *Psychological analysis of economic behavior.* New York: McGraw-Hill, 1951. Pp. ix+347. $5.00.

The general thesis of this book is that economic behavior can only be understood, predicted, and controlled by utilizing psychological methods of research and analysis. Like all social behavior, economic behavior is influenced by a variety of attitudes, habits, and motives (not solely by profit seeking); it takes into account the future ("expectations") as well as the past; it is characterized by impulsive and routine behavior in addition to genuine decision-making; and the knowledge, beliefs, attitudes, and habits that influence it are acquired and change both as a result of repetition and as a result of insights based on reorganizations of field.

The author devotes most of this book to raising psychological questions about the saving and spending decisions of consumers, the output, price, and investment decisions of businessmen and the attitudes and decisions of both consumers and businessmen under the special circumstances of inflation and wide fluctuations in business activity. Tentative answers to some of these questions are suggested on the basis of empirical evidence from surveys carried out by the Survey Research Center (and its predecessor agency).

This is an interesting and valuable book. It will be very useful for teaching purposes, and both psychologists and economists will benefit from its incisive demonstration of the gaps that exist in current economic materials—gaps which only psychological investigations are likely to fill. In the place of the good but vague intentions to "integrate psychology and economics" found spasmodically in economic literature for a generation now, this book charts in realistic detail a wide variety of specific problems, investigation of which by psychological methods promises to contribute greatly to economic understanding, prediction and control. No doubt many doctoral dissertations over the next ten years will find their starting points in this book.

To the reader whose hopes will be aroused by the title, a word of caution is needed, however. The book does not attempt a general evaluation and synthesis of research significant for the "psychological analysis of economic behavior." Rather, it is intentionally a development of an individual viewpoint, and comprises primarily a series of *suggestions for research* on the psychology of *certain kinds* of economic behavior, illustrated by analyses of *survey data* (mainly from consumers), collected by *one research organization.*

The author devotes most of his book to a stimulating inquisition of economic dogma—raising questions on the basis of psychological concepts, analogies, case studies, and survey results. Although a wealth of consumer survey data is drawn upon, the comparative paucity of other reliable psychoeconomic data today unfortunately tends to make the book seem negative, for the author generally fails to replace the demonstrably unsatisfactory with the more satisfactory. Characteristic of the book is the refrain of nearly every section that "On the basis of the studies conducted up to now, we cannot foretell in advance when the one and when the other situation is likely to occur" (p. 95), and "This may be the case under certain conditions that require further study, but need not be the case under other conditions" (p. 167). The author speaks of "the present stage of our discussions, when we are searching for a program of research" (p. 137). This results, of course, from the author's deliberate choice of the inductive method.

Many readers will question the author's selection of the kinds of psychological problems and concepts deserving greatest attention. The book abounds in references to the prevalence of habit in economic behavior except where a reorganization of field provides the preconditions for a "genuine decision." Little attention is given, however, to individual variations in *knowledge* about economic trends and market conditions, as a major factor conditioning both habitual responses and genuine decisions, or to the channels of communication through which knowledge is acquired and attitudes and expectations are influenced. There is some discussion of levels of aspiration, and the term "attitude" is often used, but how aspirations and attitudes (other than expectations) arise, how they change and how closely they may be related to action is almost entirely neglected, despite the extensive empirical work carried on by others in this field. Likewise, psychoanalytical concepts and research get short shrift. The author does not consider that existing psychoanalytic concepts are very helpful in analyzing economic behavior, or that they are likely ever to be very helpful in understanding the "regularities of behavior that are responsible for relationships between, say, income changes and saving or for cyclical fluctuations in the present American economy" (pp. 58–59). He apparently does not consider (as the reviewer does) that, for example, anxiety feelings may significantly influence consumer or business behavior (e.g., during war or in business crises) or that ego-satisfaction, anxiety-strengthened prejudices, or a search for security may have important economic results by influencing top officials in strategic companies (e.g., price leaders) in strategic industries (those whose decisions have extensive repercussions on economic development or on business fluctuations).

The author also omits certain major areas of economics. He explicitly excludes employee behavior (aptitudes, incentives, training, technological change, mobility, organizations, etc.) and the whole field of international economics (including the burgeoning new field of international investment, technological diffusion, and governmental aid programs). Moreover, although he points out that policy makers in government need surveys to help them make decisions, he does not discuss the psychological factors that may affect governmental economic activities (public services, government-owned enterprises, regulatory agencies, fiscal management, legislation on economic matters, etc.) This omission is particularly noticeable in the discussion of inflation.

Conspicuously absent in the discussion of consumer behavior is any consideration of the psychological determinants of value, of the shape of indifference curves, and of the elasticity of demand. Despite these omissions, however, the two chapters on consumer spending and saving are the strongest in the book, with significant theoretical questions raised and pertinent empirical data adduced in hard-hitting fashion.

Psychological aspects of social interaction, communication and methods of social control (advertising, education, publicity, propaganda), in their bearing upon economic behavior, are also relatively neglected in the book.

The discussion of business behavior leans too heavily, in the reviewer's opinion, on data from small enterprises. It consequently probably overstates the extent of habitual behavior in business, for small companies have fewer facilities, and perhaps less pressure, to collect pertinent information, utilize a variety of specialists, perceive and review alternatives, and otherwise engage in the processes characteristic of genuine decision-making. Moreover, it is particularly among large firms that the phenomena of price-leadership and industry-dominance occur, and that highly personal decisions may have quite significant impacts on the economy. The book also gives much less emphasis than is deserved to the relationship between business behavior and the characteristics of the industry (average size of concerns, extent of public interest, stage of growth, competitiveness, nature of product, importance of labor costs, market served, etc.) and of the concern (structure, locus of control, relative size, etc.). The discussions of business investment (drawing heavily on Terborgh's work) and of retail behavior during inflation are, however, particularly effective.

In the reviewer's opinion, the book would have gained a good deal from inclusion of empirical materials in addition to data from surveys with which the author himself has been associated. No use is made of the extensive research carried on by non-governmental organizations (other than the Survey Research Center) concerning consumer behavior and its relationship to income, geographic area, sex, race, size of town, education, reading habits, advertising stimuli, product changes, etc. There is no reference to the Psychological Corporation, for example, nor to the work of such men as Alderson, Britt, Blankenship, Link, Stanton, or Lazarsfeld (except in methodological notes). In discussing consumer and business attitudes and behavior during business cycles, *consumer* surveys are quoted, but no data are utilized from the surveys of *business* expectations made by *Fortune*, McGraw-Hill, the Securities Exchange Commission-Department of Commerce, and Dun and Bradstreet (though these are mentioned in the later chapter on methodology).

Similarly, many kinds of aggregate economic data might have been utilized to give point to the author's discussions of economic behavior, especially in view of the relative paucity of actual surveys of individual entrepreneurs. Data on inventory fluctuations, internal and external corporate financing, installment buying and other consumer credit, durable goods expenditures, expenditures for advertising outlays on plant and equipment, output and prices in different industries, etc., might well have been used when these subjects were discussed. None of the statistical series developed by the National Bureau of Economic Research or by the Department of Commerce is presented, however.

This book, therefore, is not nearly so general as its title would lead one to believe. In positive terms, what contribution does it make?

1. It presents what is by now a familiar, though generally unheeded, plea for more empirical studies of economic behavior, utilizing methods that will bring out the individual differences in motivations, expectations and extent of routine or habitual behavior involved. This plea is made more than usually effective by illustrations of assumptions or generalized statements made by prominent economists which are cast in doubt or even directly contradicted by actual survey data.

2. A good deal of material is presented to indicate the relative importance of different factors (past income, expected income, liquid asset holdings, price expectations, etc.), in influencing the spending and saving behavior of consumers. Although this material is in general limited to that collected by the Survey Research Center (and its predecessor agency), it is substantial in amount, of fairly wide scope, and unusually excellent in quality—both in terms of choice of pertinent subjects and in terms of survey technique.

3. A wide variety of questions are raised about the behavior of businessmen, with enough in the way of evidence and analogy to unsettle much of the complacency of economists (and of policy makers) about the simplicity of motivation and of other determinants of business behavior.

As a book that breaks through traditional ways of professional thinking and poses a multitude of new research problems of great significance, both for economists and for social psychology, this will, despite its incompleteness, be a landmark in the frontier discipline of psychoeconomics.

SAMUEL P. HAYES, JR.

United States Special Technical and Economic Mission to Indonesia.

ROHRER, JOHN H., & SHERIF, M. *Social psychology at the crossroads.* Pp. viii+437. New York: Harper, 1951. $4.00.

This book reports the five-day conference on "Social Psychology at the Crossroads" held in 1950 at the University of Oklahoma. Social psychology is made visible in a new geographic center—both by the holding of the meetings and the publication of this volume. The book itself consists of summaries by well-established researchers. Each chapter relates a segment of concepts and theories which the contributor feels has relevance to the development of social psychology. But the volume contains no news of significant, hitherto unreported, advances.

To the reader looking for a summary of recent developments in fields related to social psychology, this volume will be useful. The materials presented are now available only in scattered sources. One familiar with the literature, however, could predict what most of the speakers talked about and what they would have to say. MacLeod, of course, wrote about phenomenology. Herskovits treated cultural vs. psychological reality. Barker and Wright illustrated the methods they use in "psychological ecology." Arensberg wrote about interactionism. And Harry Harlow, in talking about social learning in primates, ended with a plug for "learning set" theory. The contributions, even though often repetitious of each author's earlier writings, present a valuable overview of concepts and theory.

One expects variation in quality of contribution among sixteen contributors in a conference. This book meets that expectation. Some of the reports are carefully drawn documents, summarizing in an explicit and dramatic way the recent thinking and work of their authors. Postman's forging of a general theory of cognition and Schneirla's expansion of his "levels" concept of social systems are examples of this kind of document. Other contributors were apparently caught within the press of duties and wrote merely "off-the-cuff" reports.

Only four of the sixteen contributions were on social psychology per se. The other twelve were concerned with related disciplines. The volume is therefore appropriately entitled, "Social Psychology *at* the Crossroads" (italics mine). Rohrer and Sherif exhibit the lines of cross-traffic which are influencing social psychology, covering disciplines from genetics through anthropology. There is breadth in the coverage. For instance, one of the comparative psychologists (Schneirla) treats "individual properties which promote group formation." A few chapters later, an applied sociologist (Whyte) stresses the relevance of hierarchical organizations to the formulation of a social psychology.

There was duplication—and even triplication—within the conference which seems unwarranted, especially when one notes important omissions in making the junction at the crossroads cosmopolitan. For instance, the psychiatrists were absent, even though many of them become increasingly interested in the role of the patient's social milieu. Perhaps Stanton and Schwartz, who are making observational studies of communication and interaction in hospital ward situations, would have been worth including.

There was not much effort devoted by the contributors to the particular ways in which their various disciplines *should* influence the development of social psychology. The comparative and experimental psychologists were more aware of the difference between development of their basic concepts and their subsequent application in social psychology. For instance, Volkmann was quite conscious of dangers in his reckless extrapolation of his work on anchoring points in perceptual judgment as applicable to the social psychology of education. The sociologists and anthropologists, on the other hand, took "social psychology" in their stride. They were quite unconcerned about differentiating their main focus of interest from the area being claimed by the embryonic social psychologists.

The four social psychologists represented in the volume are in essential agreement with each other in staking their claim. They posit that social psychology is the systemization of "how psychological processes function under the field conditions of group life," as Newcomb put it. And living at the crossroads provides a sound habitat if Hartley, Sherif, Sargeant, and Newcomb's reports are compared with the social psychology of the 1920's, when Allport and others conceived it more narrowly in the main stream of individual psychology.

Newcomb and Sargeant emphasize in their rich, theoretical chapters the inner life of the social individual. They stress "self-concepts" and social attitudes as generated and conditioned by group roles and memberships. This concentration meant omission in the conference of another emphasis in social psychology—understanding the behavioral life of social beings in groups as stressed in the work, for example, of Bales, Carter, or Festinger, who are concerned with the description of group functioning as related to the psychological nature of man. Sherif's chapter on "A Preliminary Experimental Study of Inter-Group Relations" takes this counterbalancing position. This last chapter of the book, however, is largely a well-written recital of findings in a particular experiment and fails to counteract the theoretical emphasis of Hartley, Sargeant, and Newcomb for the volume as a whole. This omission injures the catholicity of the book.

In general, the volume ably achieves its goal of reporting the University of Oklahoma's conference on "Social Psychology at the Crossroads." The book largely consists of recapitulations of theory and facts in many fields which have bearing upon the development of social psychology. It also includes two excellent summaries of current theories in social psychology, conceived as the impact of social phenomena upon the psychological functioning of human beings.

HAROLD GUETZKOW.

Graduate School of Industrial Administration, Carnegie Institute of Technology.

QUEENER, E. L. *Introduction to social psychology.* New York: Sloane, 1951. Pp. xiv+493. $5.00.

This text is somewhat different from others in social psychology in both content and organization. The trend toward greater use of anthropological and sociological material is pushed further than is yet customary. One objective—the understanding and prediction of interpersonal behavior—determines what data are selected from these two areas and from social psychology, and how they are organized.

The author makes the plausible assumption that it would be helpful, for predicting the behavior of any one person, to know: the characteristics of (1) human beings in general and of the particular (2) culture, (3) class, (4) caste, and (5) sex to which the individual belongs, and of the particular (6) group and (7) situation in which he happens to be at the moment; also his (8) individuality (personality). Accordingly, the author devotes almost the entire book to an examination of these eight variables (the last two chapters deal with research theory and methods).

The author believes that there is a body of integrated psychological theory—field, learning, psychoanalytic, and so on—applicable to those variables. Actually, he uses mainly the first two and especially the drive reduction learning theory.

Perhaps the reviewer lacks the anthropological and sociological background to evaluate critically the author's examination of culture, class, and caste, but to him that examination seems comprehensive, meaningful, and valuable. He is less enthusiastic, however, about the analysis of groups, sex character, personality, and especially of "the human-wide variable." Queener painstakingly follows a few authors and his own thinking about these topics through many twists and turns; but in that detailed analysis he sometimes seems to overlook equally fruitful approaches. In a discussion of leadership, for instance, he follows mainly the excellent lead of Lewin and his followers, but fails to use other equally fruitful recent work on leadership. This failure leads him to conclude tentatively that there is "a common element running through all types of leadership" (p. 275).

The "human-wide variable" is tentatively identified with instinct (p. 34). Four possible human instincts—self-preservation, maternal, competitive, and heterosexual—are analyzed and rejected as human instincts. Like many another social psychologist and sociologist, the author concludes that since there are no human instincts there is no socially relevant behavior which you could predict if you knew only than an organism is human (p. 84). He does not say so, but he implies that this knowledge would not even help you in making such predictions. He does admit that innate determinants of human behavior are not entirely lacking (p. 84), but he does not elaborate upon that statement other than to indicate (pp. 74–78) that it may be inherently easier for human beings to learn certain behavior patterns than their opposites, wiz., competition than cooperation.

There is little or no reference to human-wide variables which are the product of inevitable human learning in all environments, irrespective of culture (Dennis' autogenous learning). The author seems to be horrified, in fact, that anything might be inevitable about human nature. He makes no mention either of the universals found in all cultures, like speaking some language. In part, these omissions are deliberate; but, in part too, they seem to stem from overlooking much material of value in predicting social behavior.

This text, like most of those in social psychology, does not build on physiology or on general and child psychology in any comprehensive and thorough fashion. That was excusable in McDougall's day when general psychology was dominated by a sterile sort of introspectionism and there was no child psychology. But for the social psychologists of today to follow blindly the precedent set by McDougall and to develop their own inadequate versions of general and developmental psychology seems inexcusable. Social psychology texts need either to expand into two volumes, or, as seems wiser to your reviewer, to limit their domain.

These strictures apply, of course, to most texts in this area and not in any greater degree to this particular one than to the others. This is a systematic, well written, and richly illustrated text.

CLARENCE LEUBA.

Antioch College.

MORENO, J. L. *Sociometry, experimental method and the science of society: An approach to a new political orientation.* Beacon, N. Y.: Beacon House, 1951. Pp. xiv+220. $6.00.

Moreno's reputation among American social scientists rests primarily on two specific contributions—the systematic measurement of patterns of choice and rejection among particular groups of people, usually referred to as sociometry; and the diagnostic and therapeutic use of role playing, usually referred to as psychodrama or sociodrama. Moreno himself, however, sees these contributions as secondary rather than primary; his major achievement, he feels, is the formulation of a new science of human relations, and the term "sociometry" in the present volume is used mainly in this expanded sense.

The book is made up of 34 papers, ten of which have not previously been published. They range in time from 1923 to 1951 and in scope from very brief notes to long theoretical discussions. The best account of Moreno's general theoretical position is the concluding article, entitled "Origins and foundations of interpersonal theory, sociometry and microsociology." In this he states: "Sociometry aspires to be science within its own right. It is the indispensable prologue and preparatory science for all the social sciences. . . . It is not merely a slogan indicating a special type of research, a single method or a number of techniques."

A large portion of the book is devoted to explanation of the ways in which sociometry is different from and better than field theory, Gestalt theory, psychoanalysis, psychometrics, Marxism, and traditional experimental method. A few papers describe specific empirical applications of the sociometric approach. One excellent discussion, "Psychodrama and group psychotherapy: I," presents the essential features of psychodrama in capsule form.

Several articles written within the last few years justify the book's subtitle, "An approach to a new political orientation." Here we learn that "Sociometry is the sociology of the people, by the people and for the people," and also that "The revolutions of the socialistic-marxistic type are outmoded; they failed to meet with the sociodynamics of the world situation. The next social revolution will be of the 'sociometric' type." After these brave words it is a little disappointing to find that the most recent essay, dated 1951, advocates an essentially bureaucratic approach to sociometric reconstruction through the establishment of a department of human relations within the United States federal government.

It is probably evident by now that this reviewer is unable to take seriously either the claim that sociometry is a basic science coordinate with sociology, anthropology, and psychology, or that it offers any great promise for the worldwide alleviation of social ills. For this reason the present volume seems of considerably less value than other sociometric writings reporting the results of limited therapeutic efforts and specific empirical research.

JOHN HARDING.

Commission on Community Interrelations of the American Jewish Congress.

VOLKART, EDMUND H. (Ed.) *Social behavior and personality. Contributions of W. I. Thomas to theory and social research.* New York, Social Science Research Council, 1951. Pp. ix+338. $3.00.

This volume consists of seventeen selections from the writings of W. I. Thomas. In addition, there is a thirty-two page introduction by the editor, a full bibliography of Thomas' papers and books, and a short biographical note that tells nothing of the man as a personality.

Since the present reviewer knew Thomas for thirty years, beginning as his graduate assistant in 1917, he prefers to comment on the editor's interpretations rather than to attempt more than brief notations regarding the content of the particular papers. If some of these comments seem a trifle severe, it is because this reviewer quite as much as the editor wants to see Thomas placed in proper perspective in sociology and social psychology.

Before making these comments, a word about the organization of the selections. The volume is divided into four parts.

1. The first, "Social science and social behavior," contains, among others on method, the much-discussed "Methodological Note" from *The Polish Peasant in Europe and America*, a joint work with Florian Znaniecki. It has never been made clear—either by Volkart or others—just how much of this "Note" was Thomas' own thinking and how much of it represented the interests of the co-author. Certainly the dichotomy "value" and "attitude" does not appear in other of Thomas' works and, as will be noted below, he had no interest in formal system building. Moreover, the "Note" was distinctly an afterthought and had no place in the design of the research in question.
2. The second part, "Social behavior and personal dynamics," contains, along with other papers, the classic statement on his theory of the four wishes and a suggestive piece from the *Polish Peasant*, "A theory of social personality."
3. The third part, "Social behavior and cultural dynamics," has several selections dealing with various aspects of social-cultural change. One of these, on the assimilation of immigrants, is from *Old World Traits Transplanted*, a book for which Thomas "was primarily responsible" but which was not made clear at the time it was published. (Two of his collaborators, Park and Miller, appear as the authors.) This is a recognition long overdue.
4. The final part contains an extended document "Outline for a program for the study of personality and culture." This was prepared for the Social Science Research Council and represents perhaps the first full review and appraisal of the field of culture and personality, just then emerging.

The most serious deficiency in this book is the almost complete neglect of Thomas' work prior to 1917, with the exception of one selection from his *Source Book for Social Origins* (1909). For the record it is well to indicate that the *Source Book* was one of the most important contributions of its time to sociology and social psychology. In this book, as in some of his earlier papers, Thomas made telling criticisms of the doctrine of recapitulation, of social evolutionism, and of the thesis of sharp biological differences among the races, all dogmas widely accepted at that time. Moreover, his criticism of what he aptly termed "particularism" was most appropriate. It gave clear evidence of his appreciation of multiple causation in relation to personality, society, and culture.

Another point is the editor's implication regarding Thomas' contribution to systematic theory. The editor confuses systematic theory with the making and using of hypotheses for research. As Volkart well points out, Thomas was empirically oriented and his suggestions as to hypotheses and techniques for research were excellent for a period when most social psychologists and sociologists were still given to armchair theorizing. While deeply concerned with empirical studies, Thomas never, to my knowledge, showed any interest in system building. In fact, he always seemed to be temperamentally opposed to the construction of formal systems of thought.

A longer view of Thomas' development will show that in his early years he had a strong biological orientation, as is clear from certain of the early papers reprinted in *Sex and Society* (1907). Later he came up with his theory of the four wishes which beyond the merest statement was never systematically elaborated. All he did was to provide many case histories to illustrate each of the wishes in operation. Certainly the doctrine of the four wishes showed an appreciation of the subjective, inner elements in motivation, but they were not tied together logically with other theoretical considerations of human behavior. Later Thomas took up what he called "the situational approach." Yet, as Volkart points out, he never did resolve the implicit conflict between his earlier and later views regarding the interplay of subjective and objective elements in personality. But apparently this did not bother Thomas, who was not interested in logic-tight formulations. His chief concern was to expose the rich variety of human behavior.

Thomas' great confidence in the use of personal documents as basic source data on personality is clearly set forth in various of the reprinted selections. In this connection a further point may be made. At heart Thomas was an artist. He began his academic career in the field of literature, and it was from this background that he developed his concern with the personal document. While he saw the personal document as a source for certain hypotheses regarding important general aspects of behavior, he always had a strong curiosity about the idiosyncratic and even bizarre in human personality. In this regard he reminds one of those psychologists who stress the unique and autonomous features of personality rather than the more universal.

In closing this review I want to say that the editor has done an excellent job in indicating in his introduction and notes the many implications of Thomas' work for contemporary students of personality and culture. It has always concerned me that social psychologists whose roots lie in academic psychology have shown so little interest in, or knowledge of, the important work of W. I. Thomas. Let us hope that the publication of this collection of papers will at least help them to appreciate his place in the history of this important field.

Northwestern University. KIMBALL YOUNG.

BOHM, EWALD. *Lehrbuch der Rorschach-Psychodiagnostik für Psychologen, Arzte und Pädagogen*. Bern: Hans Huber, 1951. Pp. xxiii+406. Swiss fr. 36.

This textbook of Rorschach's perceptanalysis for psychologists, physicians and educators was written in German by a Dane and published in Switzerland. It aims to survey the present status of perceptanalysis without making the reading of Rorschach's Psychodiagnostics superfluous. Of necessity it repeats Rorschach's main ideas to make the new developments comprehensible. The book is based mainly on Swiss literature but it also uses Scandinavian and several American sources. The author tried to cover all the literature that was accessible to him. The survey is conscientious, the presentation is clear and succinct. The book offers an excellent opportunity to acquaint oneself with most European contributions including Bohm's own. Multicolored loose tables, in a separate appendix, aid in rapidly identifying the normal and small detail areas. However, there are no statistical frequency tables (Rorschach's own were not copied) but frequencies are indicated often in approximate numbers. It is better to have approximate figures that are valid than precise numbers that might not be valid. The thick book contains numerous details but they are well organized and cluster around the leading concepts.

The discussion of the administration of the test is one of the best in the literature. Bohm sounds like an expert teacher who is aware of the errors that students are likely to commit. His review of the Rorschach as an aid in neuropsychiatric diagnosis probably is the most complete available within the covers of one volume. The author relates the Rorschach well to general psychopathology and to clinical psychiatry, and thus, so to speak, presents the Rorschach in its natural habitat. The reader is continually warned of the tentative nature of many statements. It may be that Bohm would not have been reticent about some of the problems which he presents as unsolved if he had been more familiar with the pertinent literature in this country. He would not then have limited the function of the test to that assigned to it by its creator. Applying to the human personality terms used in explaining the functions of the nervous system, Bohm states that the Rorschach method provides a description of a "centrally conditioned (located in the brain) selection of impressions." In other words, the Rorschach enables us to make an inventory of socio-psychological stimuli which the individual perceives, consciously or unconsciously. According to Bohm, as well as Rorschach, the test can not—except under rare and favorable circumstances—tell us about the efferent behavior, about the overt reactions to the perceived stimuli, about the inhibitions which might interfere with overt reactions, or about the amount of available energy.

The author's conservative attitude, inimical to the spirit of reformation, is strong and deliberate. He stands by Rorschach and rejects consistently all major attempts to modify Rorschach's ideas. He does not even try to justify his rejection beyond saying that those modifications confuse. On the other hand, he is receptive to all innovations which leave all of Rorschach's own ideas intact. He corrects only when there is overwhelming factual evidence to support the correction, e.g., he rectifies Rorschach's notion that psychotics never produce a color shock. The book is provided with two indexes, a name index, and a particularly good subject-matter index. There is also a glossary of symbols in German, American, French, and Latin. Records of 31 subjects are reproduced with the scoring, the tabulation, a succinct diagnostic discussion, a brief personality analysis and a clinical summary. In the eight-page chapter on the theoretical foundations of the Rorschach test, Bohm lists experiments and observations which would be helpful in building a theory, although they were not made with the Rorschach method in mind. He cites studies on the emotional significance of chromatic colors, investigations of free association, figure and ground experiments, etc.

This readable book is rich in content. It has the merit of not misrepresenting the views of other investigators, even when the author disagrees with them. It is a good textbook in a field fraught with uncertainties and complex in its structure. It is a good aid to self-aid.

Zygmunt A. Piotrowski.

Columbia University and New York University.

NOTICE

Dr. Wayne Dennis will become editor of the *Psychological Bulletin* as of January 1, 1953. All new manuscripts should *henceforth* be sent to him at the following address:

PSYCHOLOGICAL BULLETIN
Dr. Wayne Dennis, Editor
Department of Psychology
Brooklyn College
Brooklyn, New York

The Book Review Editor will be Dr. Edward Girden. All books and other publications intended for listing or review should be addressed to him as follows:

PSYCHOLOGICAL BULLETIN
Dr. Edward Girden, Book Review Editor
Department of Psychology
Brooklyn College
Brooklyn, New York

BOOKS AND MATERIALS RECEIVED

ABRAMSON, HAROLD A. (Ed.) *Problems of consciousness.* (Transactions of the second conference, March 19–20, 1951.) Packanack Lake, N. J.: Josiah Macy, Jr., Foundation Publ., 1951. Pp. 178. $3.25.

ADKINS, DOROTHY C., & LYERLY, S. B. *Factor analysis of reasoning tests.* Chapel Hill, N. C.: Univer. North Carolina Press, 1952. Pp. iii+122. $2.00.

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of values.* (Rev. ed.) Boston: Houghton Mifflin, 1951. Pp. 12 (text) and 15 (Manual of Directions). $.30.

BERDIE, RALPH F. *Concepts and programs of counseling.* Minneapolis: Univer. Minnesota Press, 1951. Pp. 81. $1.75.

BOARD OF EDUCATION OF THE CITY OF NEW YORK. *A summary of the evaluation of the Aetna Roadometer performance test.* Divisional Bulletin No. 2. New York: Bureau of Educational Research, Division of Tests and Measurements, August 1951. Pp. 11.

BÜHLER, CHARLOTTE, LEFEVER, D. W., KALLSTEDT, FRANCES E., & PEAK, H. M. *Development of the basic Rorschach score supplementary monograph.* Los Angeles: Rorschach Standardization Study, 1952. Pp. iv+71. $2.50.

BÜHLER, CHARLOTTE, SMITTER, FAITH, & RICHARDSON, SYBIL. *Childhood problems and the teacher.* New York: Henry Holt, 1952. Pp. xi+372. $3.75.

CASSIRER, ERNST. *The philosophy of the enlightenment.* Princeton: Princeton Univer. Press, 1951. Pp. xiii+366. $6.00.

CLEUGH, M. F. *Psychology in the service of the school.* New York: Philosophical Library, 1951. Pp. vii+183. $3.75.

EKMAN, GOSTA, HUESEN, T., JOHANSSON, G., & SANDSTROM, C. I. (Editorial Committee). *Essays in psychology dedicated to David Katz.* Stockholm: Almqvist & Wiksell Boktryckeri AB, 1951. Pp. v+283.

FENLASON, ANNE F. *Essentials in interviewing: For the interviewer offering professional services.* New York: Harper, 1952. Pp. xi+352. $4.00.

FESTINGER, LEON, & KELLEY, HAROLD H. *Changing attitudes through social contact.* Ann Arbor: Institute for Social Research, Univer. of Michigan, 1951. Pp. 83.

FISHER, BURTON R., & WITHEY, STEPHEN B. *Big business as the people see it: A study of a socio-economic institution.* Ann Arbor, Mich.: Institute for Social Research, Univer. of Michigan, 1951. Pp. iii+200. $1.50.

FLESCH, RUDOLF. *The art of clear thinking.* New York: Harper, 1951. Pp. vii+212. $2.75.

FLUGEL, J. C. *A hundred years of psychology.* New York: Macmillan, 1951. Pp. 424. $3.25.

Forrester, Gertrude. *Methods of vocational guidance.* (Revised and enlarged.) New York: D. C. Heath, 1951. Pp. x+463. $4.25.

French, Thomas M. *The integration of behavior.* Vol. 1, *Basic postulates.* Chicago: Univer. of Chicago Press, 1952. Pp. xi+272. $5.00.

Freud, Sigmund. *Psychopathology of everday life.* New York: New American Library of World Literature, 1951. Pp. 168. $.35.

Gurvitz, Milton S. *The dynamics of psychological testing.* New York: Grune & Stratton, 1951. Pp. xii+368. $6.75.

Havighurst, R. J., Eaton, W. H., Baughman, J. W., & Burgess, E. W. *The American veteran back home.* New York: Longmans, Green, 1951. Pp. vii+271. $3.50.

Hinckely, R. G., & Hermann, Lydia. *Group treatment in psychotherapy: A report of experience.* Minneapolis: Univer. of Minnesota Press, 1951. Pp. x+136. $3.00.

Horner, A. Musgrave. *Speech training: A handbook for students.* New York: Philosophical Library, 1951. Pp. 176. $3.75.

Kempthorne, Oscar. *The design and analysis of experiments.* New York: Wiley, 1952. Pp. xix+631. $8.50.

Hull, Clark L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951. Pp. viii+145. $2.75.

Humphrey, George. *Thinking: An introduction to its experimental psychology.* New York: Wiley, 1951. Pp. vii+331. $4.50.

Husen, Torsten, & Henricson, Sven-Eric. *Some principles of construction of group intelligence tests for adults: A report on construction and standardization of the Swedish Induction Test—the I-Test.* Stockholm: Almqvist & Wiksell, 1951. Pp. 100. 5 Sw. Cr.

Jackson, Edith B., & Trainham, Genevieve. *Family centered maternity and infant care.* (Supplement 1 to Problems of Infancy and Childhood, Transactions, Fourth Conference, 1950. Sponsored by Josiah Macy, Jr. Foundation, N. Y.) Caldwell, N. J.: Progress Associates, 1950. Pp. 29. 1–25 copies, 25 cents each; 26–100 copies, 15 cents each; 101 copies or more, 10 cents each.

Jacobi, Jolande. *The psychology of C. G. Jung.* New Haven: Yale Univer. Press, 1951. Pp. xx+244. $3.75.

Jeffress, Lloyd A. (Ed.) *Cerebral mechanisms in behavior,* The Hixon symposium. New York: Wiley, 1951. Pp. xiv+311. $6.50.

Joseph, Alice, & Murray, Veronica F. *Chamorros and Carolinians of Saipan.* Cambridge: Harvard Univer. Press, 1951. Pp. xi+381. $5.00.

Kaplan, Louis, & Baron, Denis. *Mental hygiene and life.* New York: Harper, 1952. Pp. xiv+422. $3.50.

Kelly, E. Lowell, & Fiske, Donald W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951. Pp. xv+311. $5.00.

KUHLEN, RAYMOND G. *The psychology of adolescent development.* New York: Harper, 1951. Pp. xvii+675. $5.00.

LOEWY, HERTA. *The retarded child: A guide for parents, teachers and social workers.* New York: Philosophical Library, 1951. Pp. 160. $3.75.

LUNDIN, WILLIAM H. *Projective techniques and psychotherapy: Examination of a therapy process through the use of the projective movement sequence and Rorschach technique.* New York: Society for Projective Techniques and Rorschach, 1951. Pp. 39. $1.50.

MACDOUGALL, CURTIS D. *Understanding public opinion: A guide for newspapermen and newspaper readers.* New York: Macmillan, 1952. Pp. xii+698. $5.00.

MCGEOCH, JOHN A., & IRION, ARTHUR L. *The psychology of human learning.* (2nd Ed., revised by Arthur L. Irion.) New York: Longmans, Green, 1952. Pp. v+596. $5.00.

MCKEAN, ELLY. *It's mine.* New York: Vanguard Press, 1951. (Pages unnumbered.) $2.00.

MOLONEY, JAMES. *The battle for mental health.* New York: Philosophical Library, 1952. Pp. x+105. $3.50.

PARSONS, TALCOTT, & SHILS, EDWARD A. (Eds.) *Toward a general theory of action.* Cambridge: Harvard Univer. Press, 1951. Pp. xi+506. $7.50.

PIÉRON, HENRI. *Vocabulaire de la psychologie.* Paris: Presses Universitaires de France. Pp. 356.

PLOSCOWE, MORRIS. *Sex and the law.* New York: Prentice-Hall, 1951. Pp. ix+310. $3.95.

POOL, ITHIEL DE SOLA. *The "Prestige Papers": A survey of their editorials.* Stanford, Calif.: Hoover Institute Studies, Stanford Univer. Press, 1952. Pp. 146. $1.75.

POWELL, NORMAN J. *Anatomy of public opinion.* New York: Prentice-Hall, 1951. Pp. xi+619. $5.50.

RAY, WILBERT S. *A laboratory manual for social psychology.* American Book Co., 1951. Pp. vii+173. $3.00.

RALLI, ELAINE P. (Ed.) *Adrenal cortex.* (Transactions of the second conference, Nov. 16–17, 1950, New York, N. Y. Sponsored by Josiah Macy, Jr. Foundation.) Caldwell, N. J.: Progress Associates, 1951. Pp. 209. $3.00.

RUNES, DAGOBERT D. *Spinoza dictionary.* New York: Philosophical Library, 1951. Pp. xiv+309. $5.00.

SANDSTROM, CARL IVAR. *Orientation in the present space.* Stockholm: Almqvist & Wiksell, 1951. Pp. 193. 18 Sw. Cr.

SARASON, SEYMOUR B. *Psychological problems in mental deficiency.* New York: Harper, 1949. Pp. x+366. $5.00.

SCHNEIDERS, ALEXANDER A. *The psychology of adolescence: A factual* *and interpretive study of the conduct and personality of youth.* Milwaukee, Wis.: Bruce Publishing Co., 1951. Pp. xii+550. $4.00.

SENN, MILTON J. E. (Ed.) *Symposium on the healthy personality.* (Transactions of Special Meetings of Conference on Infancy and Childhood, June 8–9 and July 3-4, 1950, New York, N. Y. Sponsored by Josiah Macy, Jr. Foundation.) Caldwell, N. J.: Progress Associates, 1950. Pp. 298. $2.50.

SENN, MILTON J. E. (Ed.) *Problems of infancy and childhood.* (Transactions of the fourth conference, March 6–7, 1950. Sponsored by Josiah Macy, Jr. Foundation.) Caldwell, N. J.: Progress Associates, 1951. Pp. 181. $2.25.

SHUTTLEWORTH, FRANK K. *The adolescent period: A graphic atlas and a pictorial atlas. Monogr. Soc. Res. Child Developm.* 1949, 14, Ser. No. 49, 1 & 2. Evanston: Child Development Publ., 1951. (Pages unnumbered.)

SLOTKIN, J. S. *Personality development.* New York: Harper, 1952. Pp. x+401. $4.50.

SMITH, DEAN A., & WOODRUFF, M. F. A. *Deficiency diseases in Japanese prison camps.* London: His Majesty's Stationery Office, 1951. Pp. viii+209. SD 6/6.

SMITH, F. V. *The explanation of human behaviour.* New York: Macmillan, 1952. Pp. ix+276. $2.75.

STETSON, R. H. *Motor phonetics: A study of speech movements in action.* Amsterdam: North-Holland Publishing Co. (for Oberlin College), 1951. Pp. xi+212. $3.50.

STONE, C. P. (Ed.) *Annual review of psychology.* Vol. 3, 1952. Stanford, Calif.: Annual Reviews, Inc., 1952. Pp. ix+462. $6.00.

TABA, HILDA, BRADY, ELIZABETH H., ROBINSON, JOHN T., & VICKERY, W. E. *Diagnosing human relations needs.* Washington: Amer. Council on Educ., 1951. Pp. xi+155. $1.75.

TINBERGEN, N. *The study of instinct.* New York: Oxford Univer. Press, 1951. Pp. xii+228. $7.00.

UNIVERSITY OF TORONTO (Staff of the Institute of Child Study). *Twenty-five years of child study.* Toronto, Canada: Univer. of Toronto Press, 1951. Pp. xiii+182. $4.00.

WATSON, ROBERT I. *The clinical method in psychology.* New York: Harper, 1951. Pp. xii+779. $5.00.

WILLIAMS, GARDNER. *Humanistic ethics.* New York: Philosophical Library, 1951. Pp. xii+ $3.75.

YOUNG, Pauline V. *Social treatment in probation and delinquency. Treatise and casebook for court workers, probation officers and other child welfare workers.* (2nd ed.) New York: McGraw-Hill, 1952. Pp. xxvi+536. $7.00.

L'année psychologique. (Cinquantième année. Vol. jubilaire offert en hommage à Henri Piéron.) Paris: Presses Univer. de France, 1951. Pp. xvi+718.

Vol. 49, No. 4, Part 1 July, 1952

Psychological Bulletin

THE EFFECTS OF PSYCHOLOGICAL STRESS UPON PERFORMANCE[1]

RICHARD S. LAZARUS, JAMES DEESE, AND SONIA F. OSLER
The Johns Hopkins University

An understanding of the effect of psychological stress upon skilled performance is of great theoretical and practical importance. People are often faced with the necessity of performing skilled work under conditions which are highly stressful. Such is obviously the case in military combat. The effectiveness of a pilot, gunner, or radar observer must be maintained even when he is threatened by physical injury or harassed by the need to hurry the performance of a complicated task. The obvious fact that human beings are often required to work under stress does not call for further elaboration.

The problems of stress involve questions of emotions, motivation, and learning. Many theoretical issues in these fields are of basic importance in an analysis of the effects of stress upon performance. This fact has been recognized; most of the experimental work upon stress has been undertaken for theoretical rather than practical reasons. The problem of the effects of stress cuts across many fields. We are able to draw many hypotheses concerning stress from the theoretical constructs of motivation, emotion, and learning.

We shall begin the discussion of the work in this field with an analysis of the concept of stress. Following this we shall describe the various experimental techniques used to induce stress. This aspect of the problem is important because there are many methodological difficulties inherent in the production of a genuinely stressful situation. Then we shall review, briefly, the kinds of performance which have been studied under stress. Finally, we shall discuss some of the theoretical implications of the work on stress and skilled performance.

---

[1] This research was supported in part by the United States Air Force under contract, AF 33(038)-13253, monitored by the Perceptual and Motor Skills Research Laboratory of the Air Training Command Human Resources Research Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

THE CONCEPT OF STRESS

It is not possible to discuss intelligently the work on psychological stress without dealing with the problem of the concept of stress. The definitions of stress that have been given from time to time are inadequate for several reasons. It is possible to think of stress in terms of situations. For example, we say that a crucial examination is stressful to the participants, or that combat is stressful to soldiers. One difficulty with this approach is that these situations are not reacted to uniformly by all people. We cannot predict the behavior of individuals by simply describing the situation. One person may tremble, sweat, experience discomfort, and show signs of behavioral disorganization. Another may show an impairment in performance with no other subjective concomitants. Still others may show no measurable effects from the situation.

In most of the research on stress, the experimenter selects a situation which, from past experience, seems to be threatening to most people. Implicit in this selection is the necessity of identifying stress with the motivations of the people who are being tested. However, because people differ in motivations and in the ways they deal with them, it is never really possible to define a general stress situation. The situation will be more or less stressful for the individual members of the group, and it is likely that these differences in the meaning of the situation will appear in terms of performance.

It is also possible to define stress by emphasizing the reactions or responses of an individual rather than the situation. The trouble with this approach is similar to that encountered when we emphasize the situation. What kinds of reactions should we measure? Are we to consider changes in skilled performance as the "measure" of stress, or are we to consider changes in subjective report? It is apparent that these things may change independently of one another. Moreover, these changes are a function of many unrelated variables. For example, skilled performance may be affected by a change in motivation or a change in approach to the task. It would be meaningless to identify these changes as the effects of stress.

Since stress cannot be defined in terms of stimulus or response operations alone, it is necessary to think of it in terms of an intervening variable.[2] The additional concept that is necessary is that of motivation. Stress, therefore, is really a secondary concept, built upon the relationship between a primary concept, motivation, and the situation in which motivated behavior appears. We would then think that stress occurs when a particular situation threatens the attainment of some goal. The actual responses that the individual may show will depend partly upon the kinds of mechanisms that have been previously established.

---

[2] Since the submission of this manuscript an interesting article by Brown and Farber, entitled "Emotions conceptualized as intervening variables with suggestions toward a theory of frustration," has appeared in this JOURNAL (November 1951, 48, 465–495). In it Brown and Farber offer some opinions which appear to be parallel with, and certainly related to, our theoretical discussion of psychological stress.

This viewpoint demands that the concept of motivation itself be explored. The psychologist who is interested in problems of human behavior finds it very difficult to estimate from measures of behavior the kind and degree of motivation involved in a particular situation. There is general agreement among psychologists that it is ultimately essential to do this in order to account for the enormous individual differences that are found in behavior. In studies of psychological stress individual differences tend to be one of the main findings.

Pointing out the parallel between the problems of psychological stress and those of physiological stress will illustrate the difficulty in dealing with motivation. Physiological stress does not seem to involve the same definitional problems that psychological stress does, because the "motivational component" in physiological stress is stated in terms of the well-worked-out mechanisms of homeostasis. Selye (34) has defined physiological stress as any condition that produces the "adaptation-syndrome," which is the reaction of the organism in returning to the homeostatic state.

The psychologist has no adequate way of defining the psychological condition that corresponds to the homeostatic steady state. Consequently, the use of the term *stress* must necessarily be a little looser than we would like it to be. When we speak of tension-systems, what we are really doing is postulating a psychological steady-state as a lack of tension. What needs to be investigated are the properties of such a state and deviations from it.

The solution of most experimenters who have studied the responses of groups under stress has been to produce situations which are thought to thwart the motives of most people. This is an adequate solution as long as one is not attempting to account for the reactions of any individual. If the experimenter tries to account for a particular individual's response, this assumption is not satisfactory.

EXPERIMENTAL PROCEDURES FOR PRODUCING STRESS

The principal problem in the study of behavior under stress has been the production of realistic stress situations. A variety of techniques have been tried. Indeed, it might be said that no two experimental studies in the literature exactly duplicate the same technique. This variety of method has led to considerable confusion, since it is likely that each of these techniques has a somewhat different effect upon performance. It is important to review the main techniques that have been used. These techniques fall into two main classes: (*a*) stress induced through failure, and (*b*) stress induced by the task itself.

Stress induced by failure. Failure or threat of failure at a task has been the method most frequently used in experiments on stress. This has been specifically done in the following ways:

1. By presenting the subject with an unsolvable task. In this procedure the subject must work at a task that, unknown to him, cannot be solved. There are a number of ways in which this condition may be established. For example, anagrams for which there are no solutions may be mixed with a group of solvable ones. Or, in another case, a subject may be required to retain a list of digits which is beyond any individual's memory span.
2. In another type of failure-stress, the subject may be interrupted at the task before he could possibly have finished. The task may consist of a group of arithmetic problems. Before he has finished all of the problems, he is interrupted, and told that his time is up.
3. A common technique for the production of failure-stress is the introduction of false norms which indicate failure even if performance has been adequate. The use of this technique may incorporate aspects of the first two procedures. For example, an individual may be interrupted before he has completed all of the items on an arithmetie test, and then told by the experimenter that anyone who has not finished has done poorly. The subject may be told at the outset of the experiment that anyone with normal intelligence (or anyone who expects to be successful in college, or have a decent career in the Air Force, etc.) ought to be able to complete this task within the time limit. In the latter case, failure to complete the problems automatically tells the subject that he is doing poorly.

These techniques of failure-stress present some difficulties. One of the most important experimental limitations is the problem of the control of the subject's motivation. In order for a failure situation to be stressful it is necessary for the individual to be motivated to succeed, or at least to avoid failure. While most subjects are anxious to succeed, the real problem is that all subjects are certainly not equally anxious. Some may not be motivated at all. The usual assumption is that enough of the subjects will be sufficiently involved to become stressed. Some subjects will be seriously disturbed by the threat of failure, while others may be scarcely threatened at all. Of course, the effect of stress will depend upon what the individual expects or demands of himself.

Unfortunately, there is no way of assessing with any confidence the degree of motivation of subjects. We cannot take the effect of stress on performance itself as an indicator of the intensity of motivation because the amount of motivation is not the only determiner of the extent and direction of the reaction to stress. It is certainly to be expected that in some instances strong motivation may produce better performance. On the other hand, it is also capable of producing impaired performance. For the moment, the question of motivation in failure-stress experiments is an important, unsolved problem.

A second difficulty in the failure-stress procedures is presented by the degree of realism that can be produced. In order to be genuinely stressed, subjects must be convinced that the instructions and information given by the experimenter are *bona fide.* Usually the experimenter trusts to his ingenuity and assumes that his instructions will be accepted at face value. There are very few experiments which attempt to assess, directly or indirectly, the degree to which the subjects accept the situation as genuine. There can be no doubt that in most of the experimental situations, the subjects' reactions to the experiment vary from skepticism to severe ego-involvement.

Finally, a seldom considered difficulty in the use of failure-scores to produce the stress is the extent to which a subject will use this false information to alter his mode of attack on the task at hand. Changes in behavior toward the task as a result of failure-information may sometimes have little or nothing to do with stress itself. A subject who has been doing an excellent job may be encouraged (particularly if he is adaptable) to give up a fruitful mode of attack on the grounds that it has been unsuccessful. Thorndike pointed out long ago that an experience of failure may make a subject more variable in performance on the basis of the information it gives him.

Stress induced by working conditions and the task itself. In addition to the experiments in which stress is produced through failure, pressure on the subject may be induced by manipulating the situation in various ways so as to produce excessive demands upon him. Various forms of distractions may be included in this category. Almost any strong sensory input which is extraneous to the task at hand may serve as a source of distraction. In an experimental situation it may be an electric shock, noises, or flashing lights. Distraction can also be produced by verbal disparagement of the subject's performance by the experimenter. Such disparagement may serve as a failure-stress as well as a distraction. In some experiments it is impossible to assess the role of verbal disparagement because it is introduced while the subject is performing, and is therefore both a threat of failure and a distraction.

We have suggested that the manipulation of the experimental situation along the lines of distractions may make difficult demands on the subject, and that these demands then serve as sources of stress. Moreover, some tasks themselves appear to be inherently stressful for much the same reasons. The task may require the subject to attend to too many things at a time, or to perform too many operations at once. These demands may be increased by the experimenter as in the case of rapidly pacing the subject. In some ways this type of stress is like the situation a person faces when he is learning to drive a car. The stress here is probably less a matter of failure—although in some instances it could be—and more of making too many demands at once upon the learner. In other cases there is the added possibility of the threat of personal injury or damage to the car. We would class this stress situation as one induced by the working conditions and the task itself.

We have contrasted these two basic types of stress situations, failure-stress and task-induced stress, because the problem of motivation appears to be a little different for each of them. In the task-induced stress, the motivation depends primarily upon how the subject interprets the experiment. If he sees the distraction or pacing as "something to beat," his effort level is apt to be raised. Motivation can be roughly equated for all subjects by carefully wording the initial instructions, or can be left to vary from subject to subject by ambiguous information. Consider, for example, the effect of instructions which indicate that the purpose of the experiment is to see how well the subject can hold up under distraction as compared with the introduction of distraction as a natural part of the experiment or the task so that the subjects will not react to it as something to be overcome. Despite the methodological problems raised here, the problem for the experimenter is simpler in the task-induced stress than it is in the failure-stress experiment. The failure-stress situations depend primarily on the production of a realistic threat to the subjects' self esteem or to some goal-oriented behavior. The variability of reactions to this kind of condition must be considerably greater than to the situation where stress is induced by manipulation of the working conditions or by the excessive demands of the task.

Evaluating the differences in technique. Unfortunately, we can only guess, at present, what the effects of the differences in the techniques of production of stress might be. Most of the experimental situations which have been used in the study of stress are not directly comparable with one another. It is almost impossible to generalize with confidence from the study of the effects of any one experimental stress-producing technique. One of the important problems for future research is the study of the differences in the effects of the various procedures for inducing stress.

THE MEASUREMENT OF THE EFFECTS OF STRESS

One of the chief difficulties in the design of stress experiments is concerned with the measurement of the effects of stress upon performance. It is important to measure the effects of stress unconfounded with any other variable. The problem arises because it is often very difficult to get a measure of reaction to stress which is independent of the subject's ability to learn or perform certain kinds of tasks.

For example, if the experimental problem is to study the effects of some kind of stress upon the learning of a psychomotor skill, the question arises, "How can we differentiate the subject's 'normal' rate of learning from his rate of learning as it is affected by the stress condition?" For data which deal with groups of subjects there is a simple solution to the difficulty. It is possible, through control subjects, to estimate the typical rate of learning and compare it with the rate of learning of a comparable group of subjects under some stress condition. If the experiment is properly designed, and the control group appropriately selected so that it is adequately matched with the experimental group, then we can observe the general effects of stress as a difference in the mean performance or a difference in the variability of performance between the two groups.

The tricky problem actually arises when we are concerned with the performance of a particular individual under stress. Whenever we are engaged in predicting performance under stress from personality measures, it is necessary to obtain a measure for that particular subject. This "stress score" must be measured independently of the subject's initial ability and his change in performance due to fatigue or learning. Because individuals differ so much with respect to their basic abilities and rates of change, some estimate of each subject's performance without stress is necessary before correlational techniques may be used.

It has been possible to solve this dilemma in three ways. The most obvious technique has been to test each subject twice—a pretest under standard conditions followed by a stress test. In that way, a difference score may be obtained between performance under stress and performance without stress. If this difference score correlates with the score on the first test (as it usually does) then it is necessary to correct this difference measure so that it is free from the subject's level of performance on the first test. It is a statistical fact that subjects who do well on the first test tend to go down on retest, while subjects who do badly on the first test improve on the second test. This means that difference scores cannot be used to represent stress scores without a correction for regression on the first test. When this correction has been made, the resulting score represents the effect of stress upon performance for any one subject.

Another technique for obtaining individual stress scores is to make use of plateaus in the learning curves, or to select tasks for study that show little or no change in performance with training. In the case of plateaus, if a subject has attained a point in the learning of a skill where little or no further change in performance is expected for some time, this level might be used as a reference point against which to compare his performance under stressful conditions. Or the plateau may be induced by giving a subject so much practice that he attains his maximum achievement on the task, and no further changes are expected.

The trouble with plateaus is that they are seldom very dependable. Most plateaus are illusory, or cannot be depended upon to remain steady long enough to introduce the stress condition satisfactorily. In the case of the final plateau, where subjects have been practiced to a maximum of performance, this level is often really not a true maximum, and further practice produces a spurt in performance. Furthermore, this latter technique, used by Williams (43), has the disadvantage of producing a stress situation which can only result in a decrement in performance if the subjects have reached their physiological limit.

In summary, it may be said that it is necessary to provide some kind of base line or standard of comparison against which to evaluate the uncontaminated effects of stress upon performance. This reference point must be made free of the influence of the subjects' abilities in the skill in question. To do this, the technique which appears to present least difficulty makes use of two test periods with each subject, one under stress and the other under control or standard conditions. We should point out that even this technique is by no means ideal. Difference scores suffer from the difficulty of requiring two equivalent tests for the same subject. Moreover, it may be important which is given first, the control or the stress condition.

A third approach is possible, but so impractical that we shall dismiss it quickly. If it were possible to match individuals beforehand on ability to learn or perform the skill, then one of the pair could be given to the control condition and the other to the experimental condition. Pretests actually could be used to match subjects. However, this technique has not been used because it would be necessary to match not only initial score, but also rate of learning, approach to the task, motivation, etc. And besides, a good matching is practically unattainable, because there are a great number of uncontrolled variables.

EXPERIMENTAL STUDIES OF PERFORMANCE UNDER STRESS

For convenience, the experimental studies have been classified into those which deal with verbal tasks and those which deal with perceptual-motor tasks. This division is somewhat arbitrary, but follows along more or less traditional lines. It is possible that a more rational classification may arise out of further experimental work.

The effects of stress upon verbal performance. Several experimenters have studied the effects of stress upon intelligence, primarily verbal, in children. Using nine-year-old boys, Lantz (**15**) obtained a statistically significant impairment of Stanford-Binet scores following a failure experience, but no such effect after a successful experience. An examination by Lantz of the differential effects of this failure experience upon the various subtests indicated that tasks requiring visual or rote memory were not affected, while those involving reasoning or thinking suffered a decrement.

Along similar lines, an interesting experiment by Hutt (**13**) differentiated between the effects of failure-stress upon a group of maladjusted children and a group of well-adjusted children. Two methods of administration of the Stanford-Binet were used: (*a*) an adaptive procedure, in which a failed item was followed by one upon which success was likely, and (*b*) the standard method, which begins with easy items and ends with items which will be failed. The adaptive method, which produced greater psychological support, resulted in higher average IQ's for the maladjusted group. No difference between the two testing methods was found for the well-adjusted group. This study has considerable implication for the problem of stress in ordinary psychometric procedures.

The great majority of the studies of stress and verbal performance show deterioration or impairment as the result of the experimental conditions. Alper (**2**) finds a decrement in production in a sentence-formation task as the result of failure-stress. In another experiment, stress induced by pacing and the inherent difficulty of the arithmetic problems produced a reduced rate of learning (**40**); pacing, in this case, produced failure in 19 out of every 20 trials. Impairment of digit span was obtained with fifth- and sixth-grade children when the number of digits presented was beyond the normal span for these children (**44**).

Zeller (**45**), working with nonsense syllables, found a decrement in recall and relearning following an experience of failure; he accounted for this decrement on the basis of repression of the items which subjects associated with failure. His evidence for this hypothesis was that relearning and recall improved when the knowledge of failure was removed. Zeller's study may be interpreted as basically a stress experiment. Zeller comments that the subjects' performance progressively deteriorated during the accumulation of failure.

In another failure-stress experiment in which nonsense syllables were learned and recalled, Sullivan (37) found that (*a*) success produced the most rapid learning and the most complete recall, (*b*) slowest learning and poorest recall followed failure, and (*c*) failure was more harmful to a superior group in producing impairment than success was facilitating. The reverse was true for an intellectually inferior group.

Impairment of performance on the digit-symbol test was found by Williams (43) with the use of false failure scores. Williams was primarily interested in validating the Rorschach test. He concentrated upon the prediction of performance under stress by selected Rorschach indices.

In another experiment, stress, induced by failure to complete a task within a specified time limit, produced an increase in errors and variability in time scores of multiplication problems and in the learning of nonsense syllables (21).

Postman and Bruner (30) studied recognition thresholds under stress using three-word sentences. Psychological stress produced by failure and ridicule resulted in poorer performance. Since the stress was administered during actual performance, the results may have been due, in large part, to distraction. The authors observed that during failure, the pre-recognition guesses of the subjects became extremely reckless.

Despite this array of studies showing impairment, poorer performance has not always been found to be the principal effect of stress. Initial decrement but later improvement in code-learning was obtained as the result of stress due to pacing.[3] Hurlock (12), on the other hand, found that failure results in an initial improvement and then a later decrement. Lazarus and Eriksen (16) found that the major effect of failure-stress upon the digit symbol test was to produce greater variability between subjects; some subjects improved and others showed impairment.

Finally we might note that there have been a large number of studies aimed at the analysis of selective recall or repression that have some bearing on the problem of the effects of stress (2, 6, 10, 14, 24, 31, 45).

[3] DEESE, J., & BOWEN, H. M. The effect of task-induced stress on code learning. Unpublished research, The Johns Hopkins University, 1950.

Eriksen's data (6) suggest that poor or good recall of incompleted tasks may be a function of personality variables related to defense mechanisms.

The effects of stress upon perceptual-motor performance. The general picture of the effects of stress upon perceptual-motor performance is similar to the pattern found with verbal tasks. A large number of studies have shown impairment of perceptual-motor performance under stress. Marquart (23), with the use of visual discrimination, concluded that frustration resulted in slow learning, increased rigidity, and non-adaptive behavior. McClelland and Apicella (20) used card sorting as the experimental task and found that stress induced by false failure-scores resulted in more trials before the criterion was reached. Longer reaction times for completing pictures that were flashed on a screen were obtained by Verville (41) when subjects were given unsolvable tasks prior to the measurement of reaction time. Moreover, slower reaction times were also found by Verville for subjects who were required to solve a number of difficult problems simultaneously before the test of reaction time.

Bayton and Whyte (4), who used a rate of manipulation test, found that a success-failure sequence produced poorer performance than a sequence of failure followed by success. Because they failed to include a control group, it is impossible to decide whether the differences in performance between the conditions are due to success, failure, an interaction between these, or sampling error.

Seashore and Bavelas (33) found a decrement in mental age scores on the Draw-a-Man test under conditions of frustration. The subjects were children of varying ages. In a famous experiment designed to get at the relation between frustration and regression, Barker, Dembo, and Lewin (3) demonstrated a regression in mental age of young children as a result of stress.

One of the most intensive attempts to evaluate the use of stress as a psychometric device for personnel selection was undertaken by the Aviation Psychology Program of the Army Air Forces (27). With the use of various psychomotor tasks such as steadiness, aiming, etc., the effects of verbal threat and distraction upon performance were measured. Since this program was primarily concerned with prediction of success in training schools, very little systematic information concerning the direct effects of stress upon the measured performance is available. Many of the studies were conducted without adequate control groups. However, in our analysis of these data (27), the following facts seem to be revealed: (*a*) Where it can be established that stress produced an effect, the effect seemed to be a small decrement in performance. (*b*) An adaptation-effect to the stress condition is suggested by the fact that the stress tests were the only psychomotor tests which produced different results according to their position within the batteries of tests. The scores for the stress tests were poorest when they occurred first in the battery.

In psychomotor performance, as well as in verbal performance, impairment of ability has not always been found to be the only result of stress. Lindsley (**18**) found that stress induced by rapid pacing caused the subjects to attempt more problems, which resulted in a greater number of errors and more variability between subjects. Total scores remained approximately the same because the increased speed was offset by an increase in the number of errors. Further data in the type of experiment described by Lindsley were obtained by McKinney *et al.* (**22**). Two kinds of reactions to stress were found: most subjects speeded up performance at the expense of accuracy and were more variable, while a few subjects showed a stable performance without any increase of speed or errors. As a whole, the stressed subjects showed a reduction of efficiency in performance in terms of the percentage of correct items to the number attempted. Clark and McClelland[4] reported that subjects with mild fear of failure performed better, particularly towards the end of a test. Finally, improved performance under failure-stress was found by Gates and Rissland (**9**) with the use of motor coordination and color naming tasks.

The effects of stress upon the components of performance. Most of the studies of stress report findings concerning the total scores on the performance of some task. There is usually no attempt made to analyze these total scores into their various components and to observe the effects of stress upon them. Two subjects could obtain identical scores on a task as the result of very different patterns of performance. This is especially true where scores may be analyzed in terms of speed and accuracy and where an increase in speed may go along with a decrease in accuracy. A report in terms of total scores may obscure what is really happening to the subject's performance.

Only two studies to date have presented data relevant to this problem. In both of these studies an increase in errors was accompanied by an increase in speed as a result of stress (**16**, **18**). In one of these studies, however, speed accounted for so much of the variance of the total scores that any increase in errors was ineffective in changing the total score (16). It is probable that the effects of stress upon such components as speed and accuracy may be very different from task to task.

---

[4] Clark, R. A., & McClelland, D. C. A factor analytic integration of imaginative, performance and case study measures of the need for achievement. Unpublished research, cited in McClelland (**19**).

Adaptability to the experimental conditions and various other less easily measured aspects of performance may also be differentially affected by stress. One problem in this area is the difficulty of determining which measure represents the subject's efficiency in any particular task. In some situations accuracy may be of greater importance than speed, while in others rapidity of performance may be the important feature.

Qualitative observations of performance under stress. In connection with experimental studies of the effects of stress upon performance, many qualitative changes in behavior have been noted. Stereotyped responses (23), inattention (44), disorganized activity (36), and increased overt activity (8) have been some of the aspects of behavior under stress that have been observed. Many experimenters have reported signs of emotional upset such as sweating, tremor, subjective anxiety, pulse changes, etc. A whole new theory of the physiology of the organism (34) has grown up around the organic changes observed as the result of physiological and psychological stress.

Investigators have attempted to make field observations on the effects of stress upon performance by means of questionnaires, interviews, and rating scales. These studies have the merit that they attempt to discover the effects of non-laboratory stresses upon behavior. However, they have the great disadvantage of depending entirely upon self-report techniques. The most serious difficulty with these studies is that they do not include actual, objective measurements of performance. The questionnaires are always administered after the stressful situation has passed away, sometimes years later (5).

Dollard (5) administered a series of questionnaires on fear in battle to veterans of the Abraham Lincoln Brigade of the Spanish Civil War. These questionnaires suffered from too many limitations to be of much use. The techniques of sampling, question-wording, the nature of the attitudes measured, etc., make this study difficult to evaluate. The data do tell us something about the conditions which produce stress in actual combat and some of the personal reactions to these stresses. However, there is no way of knowing how the intense fears reported by Dollard's subjects affected their performance.

A somewhat more satisfactory questionnaire approach was attempted by Shaffer (35), with the use of veteran AAF flying personnel. Following an analysis of the results of his study, Shaffer (35) concluded that the adequate stimulus for fear is a highly motivated situation towards which the individual has no adequate means of adjustment. This definition seems to fit most of the experimental studies of behavior under stress.

Shaffer reported many different signs of emotional upset in combat. Some of his data suggested that the severe stress of battle may, indeed, have a serious detrimental effect upon the various skills required in combat. Subjective feelings of fear increased with the number of missions that were flown, though the pilots questioned believed that this was probably not accompanied by an increase in the deterioration of behavior. Somewhat greater faith can be placed in the information from this report because of the cross-checks, the closer proximity in time of the interviews to the experience in question, and the somewhat more adequate control of sampling. Nevertheless, Shaffer's study still suffers from a complete dependence upon self-report.

The studies of Shaffer are the only field observations which actually attempted to assess the effects of combat stress on performance through self-report. The statement of a majority of the men questioned was that fear increased their efficiency or at worst made no difference in it. Only from 17 to 31 per cent of the men in the various groups interviewed reported that fear decreased the efficiency of their performance. Those men who reported an increased efficiency as the result of stress said that mild fear increased their efficiency more than intense fear.

The findings of this study are extremely interesting and provocative, but there is a distinct probability that the reports of the men interviewed do not reflect the actual state of their performance under stress. It is likely that the interviewees confused effort with performance. Also, under the circumstances of testing, men might be extremely unwilling to admit behavioral disorganization under combat. The individuals who were interviewed had no good criteria for an evaluation of their own performance under stress. The stressful conditions themselves probably reduced the accuracy of self-estimates of performance. As a result, we suspect that the percentage of men experiencing no detrimental effects of stress was greatly overestimated. In light of the experimental studies on performance under stress, it does seem likely that some individuals will show a facilitation of performance under such conditions, but it also seems probable that these individuals will be in the minority in any randomly chosen sample of people. This interesting problem remains to be studied by more reliable methods.

With the use of narco-analytic techniques, Grinker and Spiegel (**11**) studied battle-fatigued soldiers of the last war. They suggested that a primary mechanism for psychological breakdown under combat involves strong guilt feelings and conflict over supposed inadequacy and cowardly behavior during combat. While it is provocative, this study is less relevant here because all of the subjects involved were psychiatric patients.

Personality correlates of behavior under stress. Very little information has been obtained about the relationship between various measures of personality and reaction to stress. The problem has theoretical as well as practical importance. On the one hand, while great individual differences in response to stress have been recognized, few fruitful attempts have been made to discover their nature. On the other hand, it would be most useful to be able to predict which people will be adversely affected by a stressful situation.

Most of the studies which have obtained relationships between performance under stress and measures of various aspects of personality have presented correlations which are statistically significant but are too small to be of practical value. For example, Taylor and Farber (39) found that submissive children showed a decrement in performance and an increase in variability under failure-stress, while ascendant children showed an improvement. In Hutt's (13) study, previously cited, maladjusted children showed impaired mental age scores on a stressful administration of the Stanford-Binet. Lazarus and Eriksen (16) have demonstrated that students with high grade-point averages in college tend to improve under stress while poor students show a decrement and greater variability in performance when ACE scores are held constant. The authors account for this relationship by the suggestion that some poor students may obtain poor grades because of the stressful nature of college examinations.

In a study that is difficult to interpret, Meadow (25) found that women with low dominance feelings were more emotionally disturbed by failure than women with high dominance. In this experiment emotional disturbance was measured by subjective report; the author found that women emotionally disturbed by the failure showed poorer performance in arithmetic and memory tests for emotional words than did those women who reported no emotional disturbance. It is impossible to tell to what extent the subjects' reports of disturbance may have been influenced by their actual level of performance. Individuals who showed poor performance may have attempted to rationalize this by claiming to be emotionally upset during the test. The study, however, is in accord with the results found by Taylor and Farber (39).

The only study that has reported relationships between performance under stress and personality variables of sufficient magnitude to be of practical use is that of Williams (43). In a validation study of certain variables on the Rorschach test, Williams reported a multiple correlation coefficient of .824 between average decrement under stress and two scoring categories on the Rorschach. There are several features about this study, however, which suggest caution in accepting the results. In the first place, the number of cases upon which the correlations were based was only 25. There are strong possibilities of sampling errors with so few cases. In the second place, the material used was overlearned, so that the scores for the subjects under stress could only decrease. Thus, nothing but "decrement" scores were obtained. In addition, it is to be noted that the stress was produced partly by the distracting influence of the threat of electric shock.

The doubts concerning the generality of the results of this study are supported by negative findings in a study correlating a large series of Rorschach variables with the effects of stress upon performance (7).

Also on the negative side is a study by Adams (1), who found no relationship between neurotic tendency scores on the Bernreuter Personality Inventory and performance under failure-stress. In a similar experiment using the Bernreuter, Marquart (23) also obtained negative results.

McKinney *et al.* (22) found no relationship between attitudes and feelings expressed on a questionnaire and efficiency of performance under stress. Some slight positive relationships are suggested, however, with variability in a few Rorschach responses. However, the measure of variability of these Rorschach responses is not described in the experimental report.

Relevant to the problem of the personality correlates of performance under stress is an experiment by Taylor (38). She found that individuals who show high anxiety as measured by a questionnaire are people who learn a conditioned eyelid response most rapidly. Since the conditioned eyelid response is of the avoidance type, the author concludes that this superiority is the result of a high anxiety drive necessary for rapid avoidance conditioning.

Performance under stress as a predictor. There have been two systematic attempts to use performance under stress as a predictor of success in other dangerous and stressful activity. One was the well-known OSS study, and the other was an attempt by the Aviation Psychology Program to design stress tests.

The more elaborate of these experiments was the program conducted by the Office of Strategic Services to select men who would be most effective in intelligence operations in enemy territory during World War II (46). The basic procedure was to subject a group of applicants to a series of stress tests designed by the psychological staff of the OSS. The test situations were highly realistic in the sense that they did a good job of simulating stressful situations to which the candidate was liable to be subjected while on duty. The men were aware that they were being evaluated. The stress that was created was basically a failure-stress. All measurements of performance were based upon ratings by continually observing psychologists. While large individual differences were found by the raters, these were not successfully related to later performance in the field. A possible reason for this was that the various subjects in this experimental program served in different theaters of war and performed different functions, so that the criterion of performance was not comparable. The OSS study cannot be considered to have achieved successful prediction of performance under stress, despite its elaborate design.

The Aviation Psychology Program (27) made an attempt to relate performance on a series of psychomotor tests given under stressful conditions to success in pilot training, bombardier training, and navigator training. No significant relationships were found between any one of the five tests used and any available measure of success in training school. Several major criticisms of this study may be given. First, the criteria used were poor ones inasmuch as they were very heavily saturated with semi-academic abilities. Stress may actually lower the validity of a test when the criterion is success in school. Moreover, performance in training school may be poorly related to performance under operational conditions that involve stress. In support of this last statement is a study of the validity of selection and classification tests in the theaters of war. The conclusion was drawn that many of the most important features of combat performance were not measured by the traditional abilities tests (17). Second, the most important factors measured by these AAF stress tests may not be ability to perform well under stress but the specific motor skills called for by the task itself. No attempt was made to measure deterioration under stress, or any other variable that would yield a picture of the effects of stress independent of the individual's ability. This is a problem of experimental design that we have discussed earlier.

It should be mentioned also that attempts to use the selection and classification battery of the AAF to predict the syndrome known as "anxiety-reaction" or "operational fatigue" also resulted in failure (42).

AN EVALUATION OF THE PROBLEM OF THE EFFECTS OF STRESS

The stress experiments mentioned earlier have yielded varying results: some show only a decrement in performance, others show improvement, and still others produce both of these effects for different individuals (an increase in variability). It is important to understand why the results have been seemingly inconsistent.

The distressing fact is that few of the experiments are comparable because they employed different kinds of conditions to produce stress and different kinds of tasks. The question of the dissimilarity of the two large classes of stressful conditions, failure-stress and task-induced stress, has already been raised. No data are available at present that allow us to make any generalization about this question.

Another important problem concerning the comparability of the various experiments cited is that of individual differences and sampling. Various investigators have used children, adults, women, men, college students, and military personnel. Inasmuch as large individual differences have been found in all of the experiments, it is reasonable to suppose that some of the disagreements between these experiments are due to differences in the samples employed. All in all, the body of experimental literature on the topic of the effects of stress on performance is completely unsystematic.

An integrated theoretical picture about the effects of stress upon performance must take account of individual differences, the finding of impairment as well as improvement of performance, the influence of different situations, and the effects of different kinds and amounts of stress. All of the psychological concepts about stress have been motivational in frame of reference. For convenience, we may categorize the explanatory concepts into those which emphasize primarily the energizing and directive aspects of motivation and those which emphasize the emotional aspects of motivation.

The energizing aspect of motivation. Miller (**28**) and Wickert (**42**) have emphasized fear as a motivation or drive. Since an increase in motivation is usually accompanied by an increased output in performance, Miller and Wickert both suggest that fear, produced in a stressful situation, may actually be beneficial to performance. They cite as evidence the fact that AAF personnel reported that under the stress of combat they performed more efficiently (**42**).

This simple notion of fear as a drive may well account for the increased efficiency that one finds occasionally for some individuals under stressful conditions. Few people would argue with the proposition that motivation does increase the adequacy of performance. It is apparent that fear produced by a stressful situation has a considerable motivational component.

The difficulty arises because of the fact that, in some instances, high degrees of motivation or fear seem to produce a decrement or impairment of performance. Neither Miller nor Wickert takes this into account. It is therefore necessary to look for mechanisms that may supplement the notion of the simple, energizing function of motivation. It is possible to think of a critical point in the amount of fear, beyond which disruption occurs.

The directive aspects of motivation. In addition to producing the impetus to action and the persistence of action, motivation has a directive aspect and a terminating effect (**26**). This means that an individual will direct his efforts towards whatever operations tend to be satisfying. When the motives have been satisfied, the activity in this direction will be terminated. In extremely powerful motivational states, it is very possible that a subject's activity will be unadaptively directed. For example, Lindsley (**18**) and Lazarus and Eriksen (**16**) showed that one of the effects of a threat of failure was an increase in the speed of performance as well as an increase in the number of errors. It is quite possible that a considerable portion of the decrements in performance reported in other experiments may be due to an inappropriate increase in speed of performance.

Another illustration of unadaptive behavior as the result of stress comes from an experiment by Patrick (**29**), who found irrational and repetitive behavior in human subjects who were placed in a problem box and subjected to severe stress.

In many of the experiments in which failure-stress was used, it is possible that telling an intelligent subject that he has done poorly will force him to alter his mode of attack, so that it may be less effective. This kind of an effect could account, at least in part, for impairment found in some of the experimental studies.

An additional way in which the directive aspects of motivation could produce a decrement in performance is the arousal, by threat, of motives that are antithetical to the activities required by the task itself. For example, Rosenzweig (**32**) suggests that reactions to frustration may be need-persistent or ego-defensive. Need-persistence means that the individual will center his attention and efforts upon the frustrated need (goal-oriented behavior), and ego-defense means that the individual will be primarily concerned with the maintenance of self-esteem.

There are many instances in which activity directed toward the maintenance of self-esteem may make the effective continuance of importance of any of these factors will depend, to a large extent, upon the conditions of the situation and the type of stress involved, as much as upon the personality and past experience of the individual. In other words, ego-defensive reactions are apt to be especially prominent in situations that utilize failure as the source of the stress. On the other hand, distraction situations, which are not accompanied by threats of failure or disparagement, are less apt to make ego-defensive reactions necessary.

Essentially, this boils down to a consideration of interactions between persons and types of stress. It would be interesting to know what kind of individual develops anxiety reactions to task-induced stress. We might guess that such people are highly motivated to perform well. The successful understanding of any individual's performance under stress depends upon some way of measuring the kinds and strength of his motivations and relating them to the characteristics of the situation in which he must perform. The fulfillment of this aim is, indeed, no simple affair.

Task components and the interactions of emotion and motivation. As we have said earlier, relatively little attention has been paid to the study of the components of the total performance of persons in stressful situations. What is going on, in terms of the effects on the subject's performance, may be masked by attention to only the total score. This score is simply the end product of his attack on the problem, his degree and direction of effort, the effects of emotional disturbance, and his abilities.

A consideration of the various components of performance may well yield a clue as to the bases of the observed effects of stress upon performance. For example, periodic blocking in a continuous task could indicate some degree of emotional disturbance in one subject. These blockings might adversely affect speed and regularity and, therefore, result in an impaired performance. On the other hand, in the same situation, another subject might show a greatly increased rate of performance at the expense of accuracy, which would suggest that the energizing and directing aspects of motivation are involved. Both subjects end up with the same total score, for different reasons, and attention to the components of that score would provide some leads about what those reasons might be.

Relating the individual components of performance to possible causes of the stress effects would not be simple. In many instances the relationship may be ambiguous. For example, excess overt activity which has occasionally been reported as an effect of stress would be difficult to assign to any of these possible causes. Impairment of judgment might reflect emotional blocking or misguided effort as a result of false failure-information. Some help would be obtained, however, by studying the combination of effects, both qualitative and quantitative. Despite the difficulties inherent in the approach we have been suggesting, an attack on the source of the various stress effects and how they operate must be made before we can begin to understand the bases of the effects of stress upon skilled performance. It is our belief that the various individual components of performance can be more successfully related to personality variables than the total score.

Implications for future research. The most obvious questions that require an answer are:

1. How do the effects of psychological stress vary with the nature of the task?
2. What are the differences in effects produced by failure-stress and task-induced stress?
3. How does psychological stress affect the various individual components of total performance?
4. What is the relationship between personality variables, past history, and performance under stress?
5. What is the relation between the capacity of the subject and reaction to stress?
6. What relation do emotion and motivation have to performance under stress?

Research in the area of psychological stress must begin practically afresh. Any systematic program must take into account the difficulty of producing realistic stress situations and making effective measurements of the stress effects which are independent of the skills required by the task itself. It seems certain that a really systematic attack on the problems of psychological stress will produce some valuable answers to both theoretical and practical problems in behavior.

REFERENCES

1. Adams, C. R. Individual differences in behavior resulting from experimentally induced frustration. *J. Psychol.*, 1940, **10**, 157–176.
2. Alper, Thelma G. Memory for completed and incompleted tasks as a function of personality: An analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403–421.
3. Barker, R., Dembo, Tamara, & Lewin, K. Frustration and regression: An experiment with young children. *Univer. Ia. Stud. Child Welf.*, 1941, **18**, No. 1.
4. Bayton, J. A., & Whyte, E. C. Personality dynamics during success-failure sequence. *J. abnorm. soc. Psychol.*, 1950, **45**, 583–591.

5. DOLLARD, J. *Fear in battle.* New Haven: Institute of Human Relations, Yale Univer., 1943.

6. ERIKSEN, C. W. Defense against ego threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 231–236.

7. ERIKSEN, C. W., LAZARUS, R. S., & STRANGE, J. R. Psychological stress and its personality correlates: Part II. The Rorschach test and other personality measures. *J. Pers.*, in press.

8. GANDINE-STANTON, DOROTHY A. A study of failure. *Character & Pers.*, 1938, **6**, 321–334.

9. GATES, GEORGINA S., & RISSLAND, LOUISE Q. The effect of encouragement and discouragement upon performance. *J. educ. Psychol.*, 1923, **14**, 21–26.

10. GLIXMAN, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.

11. GRINKER, R. R., & SPIEGEL, J. P. *Men under stress.* Philadelphia: Blakiston, 1945.

12. HURLOCK, E. B. The value of praise and reproof as incentives for children. *Arch. Psychol., N. Y.*, 1924, **11**, No. 71.

13. HUTT, M. L. A clinical study of "consecutive" and "adaptive" testing with the revised Stanford-Binet. *J. consult. Psychol.*, 1947, **11**, 93–103.

14. KENDLER, TRACY S. The effect of success and failure on the recall of tasks. *J. gen. Psychol.*, 1949, **41**, 79–87.

15. LANTZ, BEATRICE. Some dynamic aspects of success and failure. *Psychol. Monogr.*, 1945, **59**, No. 1 (Whole No. 271).

16. LAZARUS, R. S., & ERIKSEN, C. W. Psychological stress and its personality correlates: Part I. The effects of failure stress upon skilled performance. *J. exp. Psychol.*, 1952, **43**, 100–105.

17. LEPLEY, W. M. (Ed.) Psychological research in the theaters of war. *AAF Aviat. Psychol. Prog. Res. Rep.* 1947, No. 17. Washington: U. S. Government Printing Office.

18. LINDSLEY, D. B. *A study of performance under speed stress.* OSRD, 1946, Publ. Bldg. No. 1838. Washington: U. S. Dept. of Commerce, 1946.

19. MCCLELLAND, D. C. *Personality.* New York: Sloane, 1951.

20. MCCLELLAND, D. C., & APICELLA, F. S. Reminiscence following experimentally induced failure. *J. exp. Psychol.*, 1947, **37**, 159–169.

21. MCKINNEY, F. Certain emotional factors in learning and efficiency. *J. gen. Psychol.*, 1933, **9**, 101–116.

22. MCKINNEY, F., STROTHER, G. B., HINES, RUTH R., & ALLEE, RUTH A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.

23. MARQUART, DOROTHY I. The pattern of punishment and its relation to abnormal fixation in adult human subjects. *J. gen. Psychol.*, 1948, **39**, 107–144.

24. MARROW, A. J. Goal tensions and recall. *J. gen. Psychol.*, 1938, **19**, 3–64.

25. MEADOW, A. A relation between dominance feeling and a classroom test situation. *J. Psychol.*, 1940, **9**, 269–274.

26. MELTON, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* New York: Macmillan, 1942. Pp. 667–686.

27. MELTON, A. W. Apparatus tests. *AAF Aviat. Psychol. Prog. Res. Rep.*, 1947, No. 4. Washington: U. S. Government Printing Office.

28. MILLER, D. Pathological reactions to threat of failure. Unpublished Ph.D. dissertation, Stanford Univer., 1948.

29. PATRICK, J. R. Studies in rational be-

havior and emotional excitement: II. The effect of emotional excitement on rational behavior of human subjects. *J. comp. Psychol.*, 1934, **18**, 153–195.

30. Postman, L., & Bruner, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314–323.
31. Rosenzweig, S. The recall of finished and unfinished tasks as affected by the purpose with which they were performed. *Psychol. Bull.*, 1933, **30**, 698.
32. Rosenzweig, S. An outline of frustration theory. In J. McV. Hunt (Ed), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Pp. 379–388.
33. Seashore, H. G., & Bavelas, A. A study of frustration in children. *J. gen. Psychol.*, 1942, **61**, 279–314.
34. Selye, H. *Stress.* Montreal, Canada: Acta, 1950.
35. Shaffer, L. F. Fear and courage in aerial combat. *J. consult. Psychol.*, 1947, **11**, 137–143.
36. Sherman, M., & Jost, H. Frustration reaction of normal and neurotic persons. *J. Psychol.*, 1942, **13**, 3–19.
37. Sullivan, Ellen B. Attitude in relation to learning. *Psychol. Monogr.*, 1927, **36**, No. 3 (Whole No. 169).
38. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
39. Taylor, Janet A., & Farber, I. E. The effect of failure and success as a function of ascendancy and submission. *Amer. Psychologist*, 1948, **3**, 361. (Abstract)
40. Thorndike, E. L., & Woodyard, Ella. The influence of the relative frequency of successes and frustrations on intellectual achievement. *J. educ. Psychol.*, 1934, **25**, 241–250.
41. Verville, E. The effect of emotional and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, **69**, 133–145.
42. Wickert, J. Psychological research on problems of redistribution. *AAF Aviat. Psychol. Prog. Res. Rep.*, 1947, No. 14. Washington: U. S. Government Printing Office.
43. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.
44. Zander, A. F. A study of experimental frustration. *Psychol. Monogr.*, 1944, **56**, No. 3 (Whole No. 256).
45. Zeller, A. F. An experimental analogue of repression: II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, **40**, 411–422.
46. OSS Assessment Staff. *Assessment of men.* New York: Rinehart, 1948.

Received November 3, 1951.

INFLUENCE OF SOCIAL CLASS ON PERSONALITY TEST RESPONSES

FRANK AULD, JR.[1]
Institute of Human Relations, Yale University

Do people of different social classes respond differently to psychological tests? Allison Davis and his co-workers have pointedly asked this question and have done extensive research to find the answer (**12**, **14**). They have shown that children of different social classes respond differently to the usual intelligence tests. Middle-class children as a group almost invariably do better on these tests than lower-class children. Not content with knowing that children from different social classes respond differently to these tests, Davis and his colleagues asked, "Are the tests fair to children from all social classes?". Their answer is: the differences between classes are not mainly the result of differences in native intelligence, but reflect a bias of the tests in favor of the middle-class child. Davis and his colleagues have shown that many of the test items deal with matters that the middle-class child has had a chance to learn but the lower-class child has had no experience with or no incentive to learn about.

Now that the work of Davis has opened up the problem of class differences on psychological tests, it is obvious that intelligence tests may not be the only kind of test that would show differences between subjects from different social classes. It is the purpose of the writer in this paper to review the literature on social class differences in responses to personality tests.

If people from different social classes do respond differently to personality tests, it would be necessary to take these differences into account whenever we use personality tests in research or in clinical practice. For example, if we compare Rorschach test responses of children seen at a children's mental hygiene clinic with Rorschach responses of normal school children, the differences between social classes may be confounded with neurotic-normal differences. If the children coming to the clinic are mostly lower-class and the school children are mostly middle-class, the differences between the two classes (if any) are mixed up with the differences between neurotics and normals.

---

[1] This research has been supported by the Institute of Human Relations. The author expresses his appreciation for this support to the Institute and its director, Dr. Mark May. His thanks are also due to Dr. John Dollard for calling attention to the importance of social class and encouraging him to study the relationship between social class and personality.

Differences on personality tests between people of different social classes could be of two kinds: meaningful or trivial. A difference in the way two social-class groups respond to a test may be called "meaningful" when it indicates a true difference on the variable supposedly measured by the test. When two groups respond differently to a test not because they differ on the variable supposedly measured by the test but because they differ on an extraneous variable, the difference may be called "trivial" or "artifactual." Klineberg (**23**, p. 103) has given an excellent example of trivial test differences:

> Some years ago a group of Chinese psychologists applied a number of standard inventories like the Bernreuter, Thurstone, and others to Chinese students and arrived at the sad conclusion that Chinese students were very neurotic. One of the investigators closed with a plea for more mental hygiene in the Chinese universities to overcome this unfortunate situation.
>
> What had happened, of course, was that the specific items in the inventory were interpreted quite differently by the Chinese. One question went like this: "Do you allow others to push ahead of you in line?" Of course all the Chinese said "yes," which apparently is marked on the neurotic side. Well, there are no such lines in China: they all gather around together. So everybody turned out to be neurotic.
>
> That is only one item, of course, though there are many others . . . .

Differences on personality tests between social classes may, like the differences between Chinese and American students cited by Klineberg, be trivial. For example, a personality test for high school students contains the question, "Are you often left out of things other kids do?" The maker of the test might assume that an answer of "yes" indicates shyness and an answer of "no" indicates sociability. But it is possible that the answer is more related to social status than to sociability. Hollingshead (**21**) has observed that the lower-class adolescent is frequently left out of school affairs only because of his inferior status. If, therefore, there are differences between the social classes in the way they respond to personality tests, the differences may be either meaningful or trivial.

DEFINITION OF SOCIAL CLASS

The term social class as used in this paper refers to a system of socially ranked groups.[2] Social anthropologists who have studied various

---

[2] W. Lloyd Warner (**42**) made the first statement of the class-caste hypothesis, and Warner and Lunt (**45**) developed methods for studying the class system of a community. One should keep in mind that the class structure is not the only status control of individual behavior. As Linton (**26**) has pointed out, there are statuses that do not involve factors of caste and class. Sex-typing and age-grading are two examples of such status controls.

American communities in New England, in the deep South, and in the Midwest have discovered that the social conditions under which persons have access to fundamental biological and social goals are determined in many ways by a system of privilege. The people in these communities are aware of the system and recognize various social ranks. They speak of "the society people," "the solid, respectable people," "the good people, but nobody," "poor but honest folk," and "people who never amount to anything—shiftless people." In the small cities that they have studied, the social anthropologists have found three major social classes, each having two subclasses (46). The members of the community rank the three classes according to prestige as low, middle, and high, and so the social anthropologists call these classes the lower, middle, and upper classes.[3]

In the studies included in our review of the literature on class differences in responses to personality tests, amount of income, education, occupation, dwelling area, other similar characteristics, and various composite scales of status have been used to measure social status. Some of the composite scales (e.g., the American Home Scale) emphasize economic differentiae—income or possessions—while others (e.g., Warner's Index of Status Characteristics) emphasize social differentiae.

Warner's Index of Status Characteristics is strongly correlated with social-class placement by the evaluated participation method ($r = +.97$); thus it is evident that the ISC measures the same kind of thing that the evaluated participation method does (46, p. 183). Scales like the Kerr-Remmers American Home Scale apparently measure a related but by no means identical variable. Finch and Hoehn (15) report a product-moment correlation of +.71 between the ISC and the American Home Scale. The scales that utilize social differentiae are better measures of social class as Warner uses the term and as we have chosen to define it in this paper.

Characteristics like amount of income or education are related to social-class placement but not identical with it. For example, Warner and his colleagues (46) report a correlation between amount of income and social-class placement (measured by evaluated participation method) of +.89.

Although none of these measures except the ISC is substantially identical with social class as defined by Warner, they are strongly enough related to social class to justify using them as rough indices of class placement. In describing each study, the present writer has carefully noted which measure of status was employed by the investigator.

---

[3] A review of the literature on differences in behavior between members of the different social classes is beyond the scope of this paper. Komarovsky and Sargent (24) have briefly reviewed this literature. For detailed information the reader is referred to the studies by Davis (5, 6, 7, 8), Davis and Dollard (9), Davis, Gardner, and Gardner (10), Davis and Havighurst (11), Dollard (13), Kinsey, Pomeroy, and Martin (22), Neugarten (31), Sutherland (41), Warner, Havighurst, and Loeb (43), and Warner, Junker, and Adams (44).

STUDIES OF THE INFLUENCE OF SOCIAL CLASS ON PERSONALITY

In our survey, we have classified the various studies according to the personality test investigated. The tests are listed alphabetically; under each test, the studies are arranged chronologically.

Bernreuter Personality Inventory. Hoffeditz (**20**) compared the Bernreuter scores of 27 men in a low-status group, 40 in a middle-status group, and 33 in a high-status group. Her measure of social status was the Barr Occupational Scale. She found appreciable class differences on the "neurotic-tendency" and "self-sufficiency" scales and smaller differences on the "dominance" scale.[4] An analysis of variance shows all these differences to be statistically significant at the .001 level. The lower-status subjects received the less favorable scores. Hoffeditz concluded that consistent differences were found between fathers in various occupational groups and that the highest occupational group was least neurotic, most self-sufficient, and most dominant.

Minitzer and Sargent (**30**) studied the relationship between family economic status and some personality traits of college students. Their measure of social status was family income, which was reported anonymously by their 120 subjects. They compared the 12 students with highest family income and the 12 with lowest family income. According to Minitzer and Sargent, "the study revealed no striking relationship between the family economic status of students and their personality traits . . . [but] economically underprivileged students tend slightly toward the maladjusted personality pattern . . . ." One should keep in mind that there were probably very few if any lower-class students in this college group (college students are predominantly middle-class). This would tend to restrict the range of the class variable and therefore to reduce the size of correlation coefficients.

Patterson (**32**) studied the relationship of Bernreuter scores to various characteristics in parents whose children were being studied by the Fels Research Institute. He reported the correlations between income and the various Bernreuter scales for a group of 88 men. The correlations were all low, ranging from .13 to .26. The largest correlations were between income and the B2-S scale (self-sufficiency), +.26; and between income and the B4-D scale (dominance), +.22. These two correlations are statistically significant at the .05 level.

---

[4] Throughout this review, differences between groups are described as "appreciable" when they are at least as large as 1 P.E. (P.E. calculated from the distribution of an unselected population) and as "small" when they fall short of this criterion.

Sumner (40) gave the Bernreuter Inventory to 205 Negro college women. Since the subjects were all college students, one may safely assume that almost all belonged to the middle class. Sumner took from the group of 205 the 51 who scored highest on the Bernreuter neurotic scale and the 51 who scored lowest on this scale and compared their scores on the Sims Score Card. The mean status score of the more neurotic group was lower than the mean status score of the less neurotic group. This difference fell short of significance at the .05 level ($p = .084$). Sumner concluded that no significant differences were found in socio-economic status between the most neurotic and the least neurotic of the Negro college women studied.

Sumner's conclusion that no difference in status has been demonstrated between the most neurotic and least neurotic of Negro college women follows from his data, but this finding does not mean that there is generally no relationship between social status and personality inventory score. The college group is a sample of very restricted range in social status. Thus there is little chance for any relationship between social class and personality test score to show up.

The Brown Personality Inventory. Brown (2), in a study standardizing his personality inventory, computed the correlation between the test scores and pupils' scores on the Sims Score Card. His subjects were 441 fourth- to ninth-grade children from schools in three Ohio cities. The correlation between the Sims score and the personality test score was −.118 ($p < .05$). Brown concluded " . . . neuroticism does not predominate in any particular social or cultural group."

From this study one would be tempted to conclude that there is only a very slight relationship between personality test score and social status. One must keep in mind, however, that the measure of social status was the Sims Score Card, which emphasizes economic rather than social aspects of status; a measure stressing "social" aspects might have given a different result. Furthermore, it is not possible to tell from Brown's published article whether a wide range of social status was represented by his subjects. If the range was very restricted, no appreciable correlations could be expected.

Two years later Brown (3) published a further study of cultural influences on test responses. The subjects included 100 high-status, 100 middle-status, and 100 low-status pupils from Columbus, Ohio schools. These children were from 9 to 15 years old. Status was estimated from the type of neighborhood in which the children lived. On all five parts of Brown's test the low-status children differed significantly at the .01 level from the middle- and high-status children. The differences between the social classes were most pronounced on the physical symptoms, insecurity, and irritability scales. These differences were, however, relatively small. Brown concluded, "It is interesting to observe that scores in general appear to bear a close relationship to the socioeconomic status of the groups" (**3**, p. 331).

This result seems contradictory to Brown's earlier finding. The apparent contradiction can be explained by the different kind of measure of social status (a social rather than an economic measure) and perhaps by a wider range of social status among the pupils in the later study, which thus allows any differences to be shown more strongly.

Springer (**36**), troubled by the apparent discrepancy between the findings of Brown's earlier study and of his later study, investigated the relationship between social status and test scores in 327 low-status boys and girls and 473 middle-status boys and girls. The subjects were 9- to 15-year-old New York City school children. Springer used two different methods of measuring social class. First, pupils were assigned to a low-status group if they lived in a poor neighborhood, and to a middle-status group if they lived in a good neighborhood. Second, pupils' fathers were rated on the Barr scale of occupational status. Springer reported significantly greater mean neurotic scores for the low-status group than for the high-status group (that is, for the children living in the poor neighborhood than for those living in the good neighborhood), but he did not find any appreciable correlation within each status group between the Brown neurotic score and the Barr rating.

Here again it is seen that one stands little chance of finding a statistically significant relationship between test score and social status if the range of social status among the subjects is restricted. The difference in mean personality test score between the lower-class group and the middle-class group studied by Springer is appreciable.

Gough (**16**) reported a study of the relationship between social status and scores on the Brown Personality Inventory. The subjects were 73 sixth-grade children from three public schools in a Minnesota city. The three schools were selected to represent the highest, the lowest, and a middle school on the social-status continuum. There was a small, not statistically significant difference in the personality test scores of the children in the various schools. The children in the high- and medium-status schools received more favorable scores than the children in the low-status school.

Gough also used the American Home Scale to measure social status of the pupils. This scale puts a heavy emphasis on the economic aspects of status. He found only a slight correlation ($r = -.16$), not significant at the .05 level, between the American Home Scale and the personality inventory score. The higher-status children got more favorable scores.

California Test of Personality. Stott (**39**) has reported a study of 806 adolescents living in Lancaster County, Nebraska. He found that farm children got better scores on the self-adjustment part of the California Test of Personality than non-farm children ($p < .01$). Farm and non-farm children were not significantly different on the social adjustment part of the test. Among the children of non-farm families, occupational status as measured by the Minnesota occupational classification did not seem to be strongly associated with personality test score. The scores of occupational class VII, the lowest occupational group, were, however, somewhat lower than the scores of the other groups ($p < .20$).

In another paper Stott (**38**) reported correlations between the California Test of Personality and both the economic- and cultural-status level of the family and the "farm class" of the family (i.e., how good a farm the family owned or rented). The product-moment correlation between the self-adjustment part of the test and economic level of family was $+.14$ ($p < .05$). Between self-adjustment and cultural level, $r = +.20$ ($p < .01$). Between self-adjustment and farm class, $C = +.31$ ($N = 183$). The correlations between social adjustment and economic level or cultural level were not significant. The coefficient of contingency between social adjustment and farm class was $+.30$ ($N = 183$).

One may conclude that, with the measures of social status used by Stott, only a slight relationship was found in the group studied between scores on the California Test of Personality and social status. The self-adjustment part of the test showed a greater social-class differential than the social-adjustment part. The higher-status children got the more favorable scores on the test.

Minnesota Multiphasic Personality Inventory. Gough (**17**) gave the MMPI to 223 high school seniors (90 boys, 133 girls) from a Minnesota city of 25,000. He used the Sims Score Card as a measure of social status, taking the top 38 and bottom 38 as measured by score-card ratings as his high- and low-status groups. Gough found 34 items out of 550 on the MMPI with differences significant at the .02 level. The differentiating items were grouped by Gough into five classes: (*a*) positive literary and esthetic attitudes, (*b*) social poise and self-confidence, (*c*) denial of fears and anxieties, (*d*) broad-minded, frank attitudes toward moral, religious, and sexual matters, and (*e*) dogmatic, self-righteous opinions.

The descriptions apply to the choices of the higher-status group. There were no statistically significant differences on any of the MMPI scales (taking each scale as a whole) between the high-status and the low-status groups.

One must keep in mind, of course, that there were only 38 subjects in each group. The slight differences that did appear in Gough's analysis showed the high-status group with more favorable scores.

When Gough (**18**), in a second part of this study, correlated the Sims score with the various MMPI scales, he found five scales with correlations significant at the .05 level: the *K*, *F*, *Hs*, *Pd*, and *Sc* scales. Correlations of all of these scales with social status were low, ranging from −.13 to −.21. In each instance the low-status group received the more unfavorable score. Gough concluded that the high-status pupils showed more defensiveness and reserve in regard to personal affairs and problems (*K* scale), greater conventionality (*F* scale), fewer somatic complaints (*Hs*), more satisfactory over-all adjustment (*D*), less tendency toward serious personality disturbances (*Pd*, *Sc*). As we have pointed out, these correlations are all quite low.

From Gough's description of his subjects, it seems likely that most of them belong to the middle class and very few to the lower class. In view of this restriction in the range of social-status scores, it is not surprising that Gough found no very high correlations between social status and the MMPI. He did, however, discover 34 items with a statistically significant social-class differential, and he combined these into a special MMPI scale that he calls the status (*St*) scale.

Minnesota Scale for the Survey of Opinions. Menzies (**29**) studied the relationship between responses to the Minnesota Scale (short form) and various social characteristics. His subjects were 177 men and women from 15 to 29 years old who either belonged to the so-called "cellar clubs" in a medium-sized New York city or worked for the NYA. He found in general that the poorer subjects had lower morale scores and less favorable adjustment scores. For example, the respondents whose families were on relief had a poorer adjustment score than respondents whose families were not on relief, and the difference between these two groups was appreciable. Similar differences were found between those whose fathers were unemployed and those whose fathers were employed, between the NYA workers and privately employed workers, and between NYA workers and students.

The group studied by Menzies is a rather special group, and it is hard to know exactly what population his results can be generalized to. Apparently the results of his study give some slight support to the generalization that people rating lower in social status generally receive poorer scores on personality tests.

Davidson (4), in a study of the relationship between personality and income in 102 highly intelligent 9- to 13-year-old children, found no statistically significant relationship between income and Minnesota Scale family, morale, or inferiority scores.

Pintner Aspects of Personality Inventory and Pressey X-O Test. Maddy (28), in a study of the relationship between personality and parental occupation, tested 319 sixth-grade children in Washington, D. C., whose fathers were professional men or semiskilled workers. The children whose parents were professional men got more favorable scores on the Pintner Aspects of Personality Inventory. The difference, although statistically significant, was small.

The middle-class children reported fewer worries on the Pressey X-O Test. Thirty-five was the average number of worries for the middle-class child compared to fifty for the upper-lower-class child, an appreciable difference. There were no significant differences on the "disapproval" or "interests" sections of this test. "Idiosyncracies" scores were also about the same in both groups.

Maddy concluded that there were reliable differences in personality traits between the middle-class and the upper-lower-class children. She found that the differences were greater among the girls than among the boys. Although the differences found by Maddy are statistically significant, they are relatively slight except for the "worries" section of the Pressey X-O Test.

Rorschach Method. As part of a comprehensive study of the relationship between personality and economic background, Davidson (4) gave the Rorschach test to 102 9- to 13-year-old boys and girls: 49 bright and talented children attending a New York City public school and 53 attending a New York City private school. Her subjects were almost wholly middle class, since both the public school and the private school attracted mainly middle-class children. That the children were exceptional is shown by their unusually high intelligence: the average IQ was 143.

Davidson concluded that "income is not a significant source of variation either in influencing personality pattern, or in determining the degree of adjustment or in establishing specific personality characteristics" (4, p. 108). However, some scores were correlated with income. Children of higher-income families gave more *CF* and *C* responses, got a higher sum-*C* score, and gave fewer "popular" responses than children of lower-income families.

The results of this study should not be generalized beyond the kind of group studied by Davidson, namely, a group of highly intelligent, mainly middle-class children.

Schachtel and Levi (**34**) studied a group of 50 children in a wartime day nursery and a similar group in a private nursery school, the children of prosperous middle-class parents. They reported that a larger proportion of the day-nursery (lower-class) children were "not loved" than of the nursery-school (middle-class) children, and that a larger proportion of the nursery-school group were "pseudo-loved" than of the day-nursery group. These authors do not indicate on what basis they decided to label the children's Rorschach responses as indicative of a "loved," a "not loved," or a "pseudo-loved" child. Since Schachtel and Levi do not report exactly how they carried out their study, we do not know how to evaluate it as a scientific report.

Auld (**1**) has reported a comparison of the responses of lower-class and middle-class adolescents. He used data gathered by Hertz (**19**) on 41 upper-middle-class boys and by the Gluecks (**35**) on 500 lower-class nondelinquent boys. The boys studied by Hertz were 15 years old at the time of testing; those studied by the Gluecks ranged from 9 to 17 years old ($\bar{X} = 14\text{-}6$).

Auld reported differences between the number of human-movement responses, animal-movement responses, form-color responses, color-form responses, and color responses given by the two groups. The differences in number of M responses, FC responses, CF responses, and total number of color responses were all significant when tested by the chi-square test ($p < .01$). The difference in the number of pure color (C) responses was not statistically significant, since few subjects in either group gave any of these responses. The difference in the average number of animal-movement responses given by the two groups may be statistically significant, but the sources from which the data are drawn do not give sufficiently detailed information to permit chi square to be used to test the reliability of the difference. (A t test is inappropriate because of skewness of the distributions.)

These differences cannot be accounted for by the difference in number of responses between the two groups. The mean number of responses for the middle-class group is 31, for the lower-class group, 27. If movement and color responses were figured as percentages of total responses, rather than as absolute numbers, the middle-class group would still come out higher on movement and color responses.

A study such as this, utilizing whatever data happen to be at hand, has the disadvantage that the groups may differ on other variables than social class. Furthermore, there may be some differences between the groups on account of differences in methods of administering the test. The findings of Auld should, therefore, be accepted only tentatively, as suggestive of the kind of differences that may be found in more carefully controlled studies.

Schachtel, in an appendix to the book by Glueck and Glueck (35), has reported differences similar to those reported by Auld between the responses of lower-class and middle-class adolescents to the Rorschach test. He states, "The Rorschach records in the present study tend to be more meager and less expressive than those of the metropolitan middle class. As far as we can see, this is not a function of differences in age, but of the difference in social stratum" (35, p. 363).

SRA Youth Inventory. Auld (1) has reported that there are 84 items out of the 298 on the SRA Youth Inventory on which low-income and high-income pupils respond differently. Auld utilized data presented by Remmers and Shimberg (33) in the manual for the test, using a breakdown of the responses of 2500 high school pupils in the group on which the test was standardized. The subjects had been assigned by Remmers and Shimberg to a high or a low economic group on the basis of a short questionnaire asking them whether their family had a vacuum cleaner, an electric or gas refrigerator, a bathtub or shower, a telephone, and an automobile, whether they had paid lessons in dancing, art, music, and the like, and whether their father had finished high school. This short questionnaire obviously emphasizes economic aspects of status. Auld had assumed, however, that there would be more middle-class pupils in the high group than in the low group and more lower-class children in the low group than in the high group. He found 84 items (out of 298 in the test) that showed a social class differential significant at the .01 level of confidence (by pure chance one would get only 3 items out of 298 statistically significant at the .01 level). Yet the mean scores of the low and the high groups differed only slightly.

The differences found were not confined to questions that, on the face of it, were more "social" than "personal." Many of the items that the authors of the test believed indicate basic personality difficulties showed differences between the two groups. Of 101 "basic difficulty" items, 22 showed a social-class differential.

The lower-class group checked more items on seven of the eight subscales of the test. The middle-class group checked more in the eighth area. Since the checks are considered indicative of maladjustment, the lower-class group got a less favorable total score on the test.

Thematic Apperception Test (Thompson Modification). Korchin, Mitchell, and Meltzoff (25) reported that middle-class subjects gave longer stories on the TAT than lower-class subjects ($p<.01$). Their subjects were 80 middle-class and 80 lower-class men and women (20–40 years old). Half of each group were Negro and half were white.

Vineland Social Maturity Scale. Louttit and Watson (27) administered the Vineland scale to 120 entering first-grade children who acted as their own informants. Social class was measured by the kind of school the child attended (in a poor, medium, or good area) and by the father's occupation. Children attending the better schools and children whose fathers were in higher-status occupations got appreciably better social maturity scores. The authors concluded that social maturity scores are probably affected by the socioeconomic cultures from which the children come or by the parental intelligence associated with the socioeconomic culture.

Wisconsin Scale of Personality Traits. Stagner, after studying the test scores of 128 college students, concluded that "there is evidence that children reared in homes which are lower economically . . . tend to develop traits of nervousness or emotionality, introversion, inferiority feeling and social passivity or seclusiveness" (37, p. 552). The differences between the various economic groups were, however, small. Only on the *W* scale (emotional instability) was the difference between the two extreme groups as large as 20 percentile points. The college group is, of course, one with a very restricted range of social status. Stagner did not compute standard deviations or critical ratios for the differences. In this study, economic status was measured by students' self-ratings on a five-point scale.

Woodworth-Matthews Personal Data Sheet. Wrightstone (47) gave this test to 100 pupils in the fifth to ninth grades of the Summit, New Jersey schools. He used the Sims scale as a measure of social status. He reported a correlation of $-.52$ between the Sims scale and the Woodworth-Matthews score, indicating that higher social status was associated with more favorable scores on "emotional stability."

CONCLUSIONS

This paper has reviewed the studies on social-class differences in responses to personality tests. Most of the studies included in this survey show social-class differences on the tests investigated. In about a third of the studies these differences are appreciable. The differences reported are generally larger when "social" rather than "economic" measures of social status are used and when the subjects have a considerable spread in social status. When an economic measure of status is used or the group studied has a restricted range of social status, the investigators have usually reported only slight relationships or none at all between social status and the test score. On some tests the differences between classes are appreciable, on other tests the differences are quite small. Even when the differences in total score are relatively slight, however, there are often a considerable number of test items that show a marked social-class differential.

On the basis of the studies reviewed, it appears that middle-class and lower-class subjects differ in their responses to some personality questionnaires and to the Rorschach test and Thematic Apperception Test. In every study that showed differences between classes, the middle-class subjects got more favorable scores than the lower-class subjects. The evidence presented here, however, does not indicate whether these differences are meaningful or trivial.

Because members of different social classes may differ in their responses to personality tests, it is necessary to take these differences into account whenever we make use of personality tests in research or in clinical practice. If the psychologist does not explicitly take account of these differences he is likely to assume a bias in favor of his own social class, the middle class. For example, he may use norms derived from middle-class groups, like the Hertz norms for the Rorschach, or he may judge the meaning of the items in a personality questionnaire from his middle-class point of view. To avoid these errors the psychologist must consciously take social-class differences into account.

REFERENCES

1. Auld, B. F., Jr. Cultural influences on personality test responses. Paper read at East. Psychol. Assoc., Brooklyn, March, 1951.
2. Brown, F. A psychoneurotic inventory for children between nine and fourteen years of age. *J. appl. Psychol.*, 1934, **18**, 566–577.
3. Brown, F. A comparative study of the influence of race and locale upon emotional stability of children. *J. genet. Psychol.*, 1936, **49**, 325–342.
4. Davidson, Helen H. *Personality and economic background.* New York: King's Crown Press, 1943.
5. Davis, W. A. American status systems and the socialization of the child. *Amer. sociol. Rev.*, 1941, **6**, 345–356.
6. Davis, W. A. Child training and social class. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.), *Child behavior and development; a course of representative studies.* New York: McGraw-Hill, 1943. Pp. 607–619.
7. Davis, W. A. Socialization and adolescent personality. *Yearb. Nat. Soc. Stud. Educ.*, 1944, **43**, 198–216.
8. Davis, W. A. *Social class influences upon learning.* Cambridge, Mass.: Harvard Univer. Press, 1950.
9. Davis, W. A., & Dollard, J. *Children of bondage.* Washington, D. C.: American Council on Education, 1940.

10. Davis, W. A., Gardner, B. B., & Gardner, Mary R. *Deep South.* Chicago: Univer. of Chicago Press, 1941.
11. Davis, W. A., & Havighurst, R. J. *Father of the man.* Boston: Houghton Mifflin, 1947.
12. Davis, W. A., & Havighurst, R. J. The measurement of mental systems *Sci. Mon.*, 1948, **66**, 301–316.
13. Dollard, J. *Caste and class in a Southern town.* New Haven: Yale Univer. Press, 1937.
14. Eells, K., Davis, W. A., Havighurst, R. J. Herrick, V. E., & Tyler, R. W. *Intelligence and cultural differences: a study of cultural learning and problem solving.* Chicago: Univer. of Chicago Press, 1951.
15. Finch, F. H., & Hoehn, A. J. Measuring socio-economic or cultural status: a comparison of methods. *J. soc. Psychol.*, 1951, **33**, 51–67.
16. Gough, H. G. The relationship of of socio-economic status to personality inventory and achievement test scores. *J. educ. Psychol.*, 1946, **37**, 527–540.
17. Gough, H. G. A new dimension of status: I. Development of a personality scale. *Amer. sociol. Rev.*, 1948, **13**, 401–409.
18. Gough, H. G. A new dimension of status: II. Relationship of the St scale to other variables. *Amer. sociol. Rev.*, 1948, **13**, 534–537.
19. Hertz, Marguerite R. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: I. The movement factors. *J. gen. Psychol.*, 1942, **27**, 119–188.
20. Hoffeditz, E. Louise. Family resemblances in personality traits. *J. soc. Psychol.*, 1934, **5**, 214–227.
21. Hollingshead, A. B. *Elmtown's youth.* New York: Wiley, 1949.
22. Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1948.
23. Klineberg, O. General discussion on methods. In S. S. Sargent, & Marian W. Smith (Eds.), *Culture and personality.* New York: Viking Fund, 1949. P. 103.
24. Komarovsky, Mirra, & Sargent, S. S. Research into subcultural influences upon personality. In S. S. Sargent & Marian W. Smith (Eds.), *Culture and personality.* New York: Viking Fund, 1949. Pp. 143–156.
25. Korchin, S. J., Mitchell, H. E., & Meltzoff, J. A critical evaluation of the Thompson TAT. *J. proj. Tech.*, 1950, **14**, 445–452.
26. Linton, R. Problems of status personality. In S. S. Sargent, & Marian W. Smith (Eds.), *Culture and personality.* New York: Viking Fund, 1949. Pp. 163–171.
27. Louttit, C. M., & Watson, Ruth. Vineland social maturity scores of entering first grade children. *Train. Sch. Bull.*, 1941, **38**, 133–137.
28. Maddy, Nancy R. Comparison of children's personality traits, attitudes, and intelligence with parental occupation. *Genet. Psychol. Monogr.*, 1943, **27**, 3–65.
29. Menzies, R. Minnesota-Scale morale and adjustment of young men and women as related to economic and social conditions in an urban environment. *J. Psychol.*, 1943, **15**, 115–127.
30. Minitzer, S., & Sargent, S. S. The relationship between family economic status and some personality traits of college students. *Sch. & Soc.*, 1939, **49**, 322–324.
31. Neugarten, Bernice L. Social class and friendship among school children. *Amer. J. Sociol.*, 1946, **51**, 305–313.
32. Patterson, C. H. The relationship of Bernreuter scores to parent behavior, child behavior, urban-rural

residence, and other background factors in 100 normal adult parents. *J. soc. Psychol.*, 1946, **24**, 3–50.

33. Remmers, H. H., & Shimberg, B. *Examiner manual for the SRA Youth Inventory.* Chicago: Science Research Associates, 1949.
34. Schachtel, Anna Hartoch, & Levi, Marjorie B. Chracter structure of day nursery children in wartime as seen through the Rorschach. *Amer. J. Orthopsychiat.*, 1945, **15**, 213–222.
35. Schachtel, E. G. Some notes on the use of the Rorschach test. In S. Glueck, & Eleanor Glueck, *Unraveling juvenile delinquency.* New York: Commonwealth Fund, 1950, Pp. 363–385.
36. Springer, N. N. The influence of general social status on the emotional stability of children. *J. genet. Psychol.*, 1938, **53**, 321–328.
37. Stagner, R. Economic status and personality. *Sch. & Soc.*, 1935, **42**, 551–552.
38. Stott, L. H. Family prosperity in relation to the psychological adjustments of farm folk. *Rural Sociol.*, 1945, **10**, 256–263.
39. Stott, L. H. Some environmental factors in relation to the personality adjustments of rural children. *Rural Sociol.*, 1945, **10**, 394–403.
40. Sumner, F. C. Neurotic tendency and socio-economic status of Negro college women. *J. soc. Psychol.*, 1948, **28**, 291.
41. Sutherland, R. L. *Color, class, and personality.* Washington: American Council on Education, 1942.
42. Warner, W. L. American caste and class. *Amer. J. Sociol.*, 1936, **42**, 234–237.
43. Warner, W. L., Havighurst, R. J., & Loeb, M. B. *Who shall be educated?* New York: Harpers, 1944.
44. Warner, W. L., Junker, B. H., & Adams, W. A. *Color and human nature.* Washington: American Council on Education, 1941.
45. Warner, W. L., & Lunt, P. S. *The social life of a modern community.* New Haven: Yale Univer. Press, 1941.
46. Warner, W. L., Meeker, Marchia, & Eells, K. *Social class in America.* Chicago: Science Research Associates, 1949.
47. Wrightstone, J. W. Validity of the Woodworth-Matthews Personal Data Sheet for diagnosing certain personality disorders. *J. educ. Psychol.*, 1934, **25**, 39–44.

Received September 19, 1951.

THE EVALUATION OF EXPERIMENTALLY CONTROLLED CRITERIA[1]

WILLIAM W. GRINGS
University of Southern California

Many of the most relevant objective criteria of performance in military and industrial situations are rendered comparatively useless because of their unreliability. Usually they are complex, and their variance is determined by a multitude of factors other than the performance of the individuals or groups being evaluated. As a result, correlations obtained with these criteria are low and efforts to improve upon them are often frustrating. It is the purpose of this paper to consider some of the problems of estimating or predicting the improvement which might be expected to result from the exercise of experimental control over selected sources of extraneous variation within such a criterion.

As an example of a complex and unreliable criterion, and one on which much effort has been expended, consider the average error of a bomb drop, computed as a radial error (so-called circular error) from the point of impact of the bomb to the aiming point. On the face of it, this measure is an ideal indicator of the ability of a bombardier to drop bombs. Experience, however, shows that bombardier variance makes up a relatively insignificant portion of the total observed variance of the measures, for such factors as altitude and air speed of the aircraft, turbulance of the air, abilities of the pilot to hold a course, as well as equipment variations, contribute to the magnitude of the single score.

Many efforts have been made to improve circular error measures by means of statistical and experimental controls (**10**). A typical statistical method involves the computation of average standard scores for individuals, each standard score indicating the relative position of the individual in the distribution of the dependent variable for a single extraneous variable category. A similar procedure utilizes conversion factors, usually multiplicative constants determined empirically from the circular error distribution of each extraneous variable. A third method, one of experimental control rather than statistical correction, consists of criterion or test missions in which the operation of the extraneous variables is controlled either by excluding the variable (holding it constant) or randomizing its effects with reference to individual bombardiers. It is the last method with which this report is concerned.

---

[1] This paper was accomplished under Contract Nonr 228 (02), Project NR 153-093, between the University of Southern California and the Office of Naval Research.

In situations such as this (as well as in comparable laboratory studies[2]) much time and effort can be saved through the use of appropriate estimates of the increased precision to be achieved by experimental control. Frequently a meaningful estimate can be made from a pilot study designed to predict the feasibility of experimental control. Two types of estimate are considered here. One is an estimate of what the reliability (proportion of true variance) of the criterion variable would be if certain of the extraneous variables were perfectly controlled in the sense of having all of the variance removed. The second is an empirical estimate of the minimum variance of the extraneous variables which might be expected under specifically defined operational conditions.

Of these, the first permits an estimate of an ideal state in which the criterion has been corrected for the variation contributed by noncriterion sources. The second yields a practical estimate of the proportion of the obtained criterion variance which is available for prediction (correlation) purposes. Both types of estimate are simple in principle and can be achieved in several ways—through correlation analysis or by modification of analysis of variance techniques. The latter method will be used with this example.

In the simplest case of bombing error determined by only two variables, the bombardier and equipment differences, one might expect to obtain data similar to those diagrammed in Figure 1. There the hypothetical errors of 100 bombardiers flying with four different sets of equipment are tabulated. The total variance in this table may be attributed to (*a*) equipment variation, (*b*) differences among individuals, and (*c*) inconsistencies in individual's relative position on the error scale among equipment classes.[3] If each measurement of each individual with each equipment were perfectly consistent, the individual rank orders would be the same in each equipment subdivision and (assuming equal within-equipment distributions[4]) the rank differences (in circular error units) would be equal from subject to subject and from equipment to equipment. This would be the same as saying that the measurements of circular error are perfectly reliable, ignoring the variance determined by equipment and assuming no other unidentified factors present.

---

[2] The estimates involved are typical of laboratory measurement situations where the behavior of the dependent variable is complexly determined and it is desired to observe the effects of only one or two independent variables.

[3] This is to a degree a simplification, for some sources of error which could be classified and for which magnitudes could be estimated (i.e., interaction effects) have been left in the "unreliability" category.

[4] This assumption often cannot be met, for it is the nature of many extraneous sources of error to affect the variability within classes as well as the average values of the class. An example of such a variable with circular error is the bombing altitude. As altitude increases the mean error increases as does also the variance of the "within-altitudes" distribution.

The approach to such an estimate of reliability (independent of equipment effects) may be visualized in a different way. That is to "correct" for the variation determined by equipment differences. Assuming homogeneity of variance in equipment subgroups, this can be accomplished by adding to all measures in a given subdistribution the difference between the mean of that distribution and some reference point on the scale (say the mean of the middle or typical equipment distribution). This reduces the variability between equipment classes to zero and leaves only variability determined by "true" individual differences and "error" or unreliability.

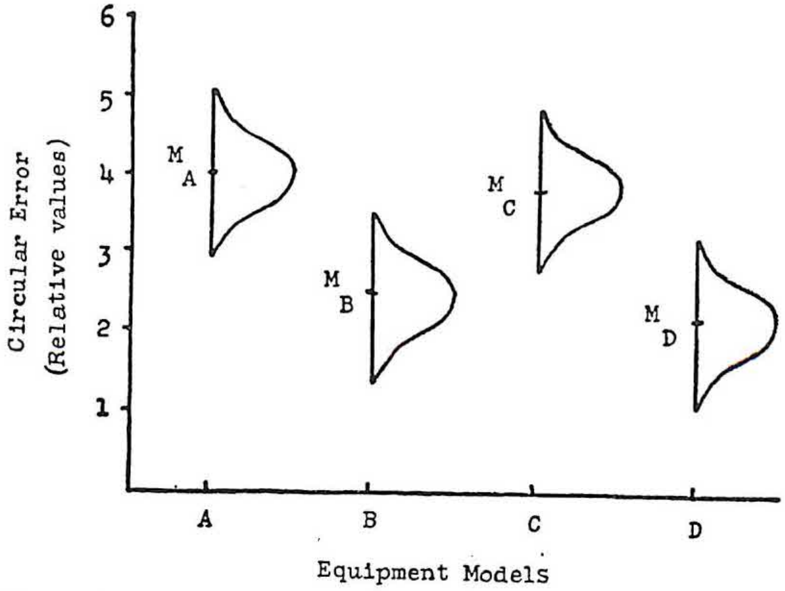


FIGURE 1. FREQUENCY DISTRIBUTIONS OF CIRCULAR ERROR SCORES FOR FOUR EQUIPMENT SUBGROUPS. (Hypothetical data; same 100 individuals on all equipments.)

The second "correction" can then be introduced into the data by removing individual differences by adding a constant to all scores in order to make individual means equal. To assume that such a process of removing individual differences by a correction based on mean differences removes individual contributions to total variability requires an assumption of homogeneity of variance within individuals—a hard assumption in some situations to satisfy. If, however, this assumption can be met (as it can in this example) and there is zero correlation of within persons variations and within machines variations, the variance remaining within the table after such a double correction is that which best indicates the "unreliability" of the measurements of the dependent variable. It may be used in the conventional reliability coefficient equation, in which r_{11} is a reliability coefficient (the equivalent of the square of the

$$r_{11} = 1 - \frac{\sigma^2 \text{ error}}{\sigma^2 \text{ total}} \qquad [1$$

eta coefficient of correlation) and the variances are those defined above.

Computationally, these estimates can be determined readily by numerous means. The simplest involves the modification of the methods of analysis of variance, since the reasoning used throughout follows that technique as it applies to the analysis of total variance into components. The chief modification required is in the degrees of freedom involved; for each variance component in analysis of variance is an estimate of population variance in the observed variable, while each variance component desired here is an estimate of the variance contributed to the total observational situation by the variable in question.

For classification schemes of this level of complexity (one control variable) estimates achieved in this fashion are identical with results secured by the use of simple conversion tables or within-class correlations. However, in classification schemes of greater complexity, particularly those of factorial design, improved estimates are possible. The estimate of error which is chosen as the best indication of unreliability may differ with the nature of the predictions. Frequently an estimate derived from an interaction sum of squares provides an appropriate indication of the likely precision in a given prediction situation.

An example of this latter type arose in attempting to make group training predictions from "learning curves" of circular error performance (4). There learning was, of necessity, inferred from a temporal (day-by-day or mission-by-mission) change in criterial (group average circular error) performance. During the observation of extraneous variables influencing criterial scores it became apparent that some of these also showed a temporal course.[5] They produced psuedo-learning effects. A test of whether this psuedo-learning effect is sufficient to mask "true" learning effects can be made by an analysis of variance significance test of the day-to-day variance over the interaction variance of equipment and days. If an estimate is then desired of the extent to which this interaction of equipment and "learning" variables influences learning predictions, the interaction sum of squares can be appropriately weighted to indicate the proportion of the total variance which is involved in this interaction "unreliability."

---

[5] This is common, for example, with equipment variations. Frequently equipment breakdown is preceded by periods of gradual change in machine function which appear in training records as "learning" variance.

It becomes apparent at this point that the specification of the variance designated σ^2 total in formula [1] is also appropriately different in different contexts. If, in the example cited, the desire had been to make individual rather than group predictions, an appropriate total variance estimate would be that determined by the variability of the individual's means—an among-individuals variance.

All of these estimates are subject to at least two general types of disadvantages. For one thing, the assumptions underlying such subdividing of variance are sometimes hard to meet. They include the assumption of independence of subclassifications or the estimation of covariances when such independence does not exist. Similarly, the independence of means and variance is required—or, more specifically, it needs to be assumed that the subclasses of the extraneous variables for which correction is being made differ in no essential aspects except their means. This is typically approached by a test of homogeneity of variance within subclasses.

In passing, it is worthy of note that these estimates do not require some of the usual assumptions of analysis of variance. Randomness of subgroups is not essential except in so far as the requirements of independence of variables is concerned, and the requirement of normality of population distribution which is basic to the *F* test is not involved here. In these respects the similarity of this reliability estimate to the eta correlation coefficient is greater than it is to the population variance estimate required for analysis of variance significance tests. The correlation of such a "corrected" variable with a second variable (say a test) is thus still a "within classes" correlation.

Another weakness in some estimates arises from the fact that they do not enable one to define the extended "practical" range of prediction (validity) coefficients made possible by the experimental controls, although they do indicate the maximum theoretical or ideal values which would occur if perfect control were achieved. Since perfect experimental control is seldom reached, it is essential in any given situation to compute other estimates designed to take into account the variability of the extraneous variable which remains after experimental controls have been instituted. Such an estimate requires either a series of pilot study checks to determine the effectiveness of controls, or some information (estimated from control charts and the like) of the extent to which controls are being achieved. Running estimates of reliability based on the same reasoning as the above can be used to compare the "achieved" reliability with the "ideal." The difference between the two estimates needed is essentially that the variance of the extraneous variable is excluded in one, while in the other it is left in and estimated from control charts or pilot analyses.

In conclusion, it will be noted that all of the foregoing situations are simply specific applications of the more general classes of reliability estimates in common use. Specifically, they represent application of the general formula proposed by Horst (**7**) and applied by him to test situations. The logic is essentially that of test-retest measurement situations as outlined by Guttman (**6**), Alexander (**1**), Cronbach (**3**), and Jackson (**9**). It has also been applied to laboratory measurement situations by Melton (**11**, **12**) and Chapanis (**2**).

REFERENCES

1. ALEXANDER, H. W. The estimation of reliability when several trials are available. *Psychometrika*, 1947, **12**, 79–99.
2. CHAPANIS, A. Theory and methods for analyzing errors in man-machine systems. *Ann. N. Y. Acad. Sci.*, 1951, **51**, 1179–1203.
3. CRONBACH, L. J. Test "reliability": Its meaning and determination. *Psychometrika*, 1947, **12**, 1–16.
4. GRINGS, W. W. Learning curves in bomb dropping. Unpublished Ph.D. dissertation, Univer. of Iowa, 1946.
5. GRUBBS, F. E. On estimating precision of measuring instruments and product variability. *J. Amer. Stat. Assoc.*, 1948, **43**, 243–264.
6. GUTTMAN, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, **10**, 255–282.
7. HORST, A. P. A generalized expression for the reliability of measures. *Psychometrika*, 1949, **14**, 21–32.
8. HOYT, C. Test reliability obtained by analysis of variance. *Psychometrika*, 1941, **6**, 153–160.
9. JACKSON, R. W. B., & FERGUSON, G. Studies on the reliability of tests. Univer. of Toronto *Dept. Educ. Res. Bull.*, 1941, No. 12. Pp. 132.
10. KEMP, E. H., & JOHNSON, A. P. Psychological research on bombardier training. *AAF Aviat. Psychol. Prog. Res. Rep.*, 1947, No. 9. Washington: U. S. Government Printing Office.
11. MELTON, A. W. The methodology of experimental studies of human learning and retention: I. The functions of a methodology and the available criteria for evaluating different experimental methods. *Psychol. Bull.*, 1936, **33**, 305–394.
12. MELTON, A. W. (Ed.) Apparatus tests. *AAF Aviat. Psychol. Prog. Res. Rep.*, 1949, No. 4. Washington: U. S. Government Printing Office. Pp. 1056 (especially p. 48).

Received December 6, 1951.

RELIABLE AND QUESTIONABLE SIGNIFICANCE IN A SERIES OF STATISTICAL TESTS[1]

JOSEF BROŽEK AND KENNETH TIEDE
The Laboratory of Physiological Hygiene
University of Minnesota

Not infrequently the psychologist has to carry out a large number of tests of significance. Analysis of items constituting a questionnaire may serve as an example.

A psychosomatic inventory[2] was administered to a group of normotensive and hypertensive individuals (systolic blood pressure of 140 mm. Hg and over, or diastolic blood pressure of 90 mm. Hg and over). The inventory consisted of a series of 228 items focussed on a variety of personality aspects described in the literature as being characteristic of individuals who are prone to develop high blood pressure. Each item was tested for significance of the difference between the percentage of "Yes" answers given by each of the two groups. With the 5 per cent level of the critical ratio (C.R. $\geqq 1.96$) as the criterion of significance, 24 items showed differences significant at this level or better.

Wilkinson (2) pointed out the necessity, under conditions of repeated tests of significance, for determining whether the number of "significant" items is sufficiently great to indicate a nonchance occurrence and provided a table giving the probability of obtaining n or more significant statistics by chance in a set of N tests. Values were given for N from 1 to 25, n from 1 to 4 for statistics significant at the 1 per cent level and n from 1 to 7 for the 5 per cent level of significance. The tables are useful but the N is too small for most item analyses. It is the purpose of this note to point out an approximate but satisfactory solution for conditions in which N is large.

Our task, to rephrase the problem, is to determine the probability of a chance occurrence of the compound event of obtaining n significant critical ratios ($n = 24$) in a series of N tests of significance ($N = 228$), the constituent events being independent. Each event has only two possible outcomes, i.e., the critical ratio is either significant or nonsignificant. The probability of chance occurrence of a significant difference at a single trial was set, by the criterion for statistical significance, at 5 per

---

[1] This work was supported in part by the Research Grants Division of the United States Public Health Service.

[2] LABORATORY OF PHYSIOLOGICAL HYGIENE, UNIVERSITY OF MINNESOTA. Personality differences between normotensive and moderately hypertensive individuals. II. Results obtained from a psychosomatic inventory. (To be published.)

cent ($p=0.05$); the probability of nonchance occurrence is 95 per cent ($q=1-p=0.95$); $p+q=1.00$.

For repeated trials the probabilities of the occurrence of "significant results" by chance may be determined by the expansion of the binomial $(p+q)^N$. For two trials we have

$$(p+q)^2 = p^2 + 2pq + q^2.$$

Substituting for p and q, we obtain $0.0025+0.0950+0.9025$, where the first term indicates the probability of obtaining by chance two significant results in the two trials, the second term refers to the probability of obtaining by chance one significant result, the third term gives the probability of obtaining no significant result by chance (probability of obtaining of a significant result through other factors than chance alone). As a rule the first two terms would be summated, yielding 0.0975, which indicates the probability of obtaining one or more significant results by chance in two trials. For three trials we have

$$(p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3.$$

The binomial expansion yields the probabilities of the chance occurrence of 3, 2, 1, and 0 significant results. Again we would summate to get the probabilities of obtaining one or more and two or more such results. It is apparent that with an increase in N the computational labor quickly becomes unmanageable.

The probability of obtaining by chance n or more critical ratios in series of N tests may be obtained exactly from the equation

$$\begin{aligned} P &= \sum_{x=n}^{N} \binom{N}{x} p^x q^{N-x} \\ &= 1 - \sum_{x=0}^{n-1} \binom{N}{x} p^x q^{N-x}, \end{aligned}$$

where $x=0, 1, 2, \cdots, n-1$. Then, in our specific case,

$$P = 1 - \sum_{x=0}^{23} \binom{228}{x} (.05)^x (.95)^{228-x}.$$

It is obvious that the direct calculation of the P value is hopelessly arduous.

Given a large N and $Np \geqq 5$ (with a $p \leqq 0.5$), the binomial distribution can be approximated satisfactorily by the normal distribution with mean $m=Np$ and standard deviation $\sigma=\sqrt{Npq}$ (**1**, p. 41). To solve our problem we compute a critical ratio, $C.R. = (n - m)/\sigma$ and determine from the available tables the area to the right of the obtained $C.R.$ value.

For our case $m = 228 \times 0.05 = 11.40$,

$$\sigma = \sqrt{228(0.05)(0.95)} = \sqrt{10.8300} = 3.29;$$

$$C.R. = (24 - 11.40)/3.29 = 3.83,$$

yielding a negligible probability ($P = 0.00006$) of obtaining 24 or more significant critical ratios in a series of 228 items tested for significance of the difference. The normotensive and the hypertensive subjects differ reliably in their responses to a psychosomatic inventory.

The binomial frequency distribution is discrete, not continuous. In carrying out the tests based on normal distribution, which is continuous, we may consider frequency 24 as occupying the interval from 23.5 to 24.5. It is more exact to calculate the area under the normal distribution to the right of the value 23.5.

It is hoped that this note will be found useful by others who are confronted with similar problems, frequently emerging when a large number of statistics is calculated. Strangely enough, the matter has not been adequately and specifically discussed in the standard texts of statistics which are likely to be consulted by psychologists.

REFERENCES

1. Hoel, P. G. *Introduction to mathematical statistics.* New York: Wiley, 1947.
2. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, 48, 156–158.

Received January 5, 1952.

THE EFFECTIVENESS OF SHORT METHODS FOR CALCULATING CORRELATION COEFFICIENTS

JOHN C. FLANAGAN
American Institute for Research
University of Pittsburgh

Two factors in recent years have emphasized the value of short methods for calculating correlation coefficients. One of these is the tendency to use larger numbers of variables. If there are 5 variables for which intercorrelations are to be obtained, the number of correlations to be calculated is 10. If there are 10 variables this number jumps to 45. If the number of variables is 20 there will be 190 intercorrelations. And if the number of variables is 100, the number of correlations to be calculated is 4950. Thus, unless electrical or mechanical equipment is available the large numbers introduce serious problems. Even if computing equipment and trained personnel are available the costs make shortcuts desirable. The second factor is the realization that large samples are essential if stable findings are to be hoped for in correlational studies. If the problem is to establish that a relationship of useful size does or does not exist between two variables, the sample may not have to be very large. To obtain predictive weights as by multiple correlation procedures, several hundred, if not several thouand, cases are likely to be necessary. For factorial analysis procedures large samples are similarly essential.

For these reasons short methods have become increasingly desirable but data on the comparative accuracy of the various short procedures have not been available. It is the purpose of this paper to discuss, first, some of the principles to be used in selecting a suitable short method for a particular problem; second, to present some data on the time required to compute coefficients by the various methods; and third, to present data regarding the relative accuracy of these procedures as compared with the longer product-moment computational methods.

TYPES OF SHORT METHODS

Turning first to the considerations involved in developing or selecting a short computational method it is apparent that shortcuts can be obtained in four ways. First, it may be possible to cut the number of operations that have to be performed on the data; second, time may be saved by substituting simple numbers for larger ones; third, simple operations such as looking up a value in a table may be substituted for more complex operations; and fourth, the operations can be performed only on selected data or only once for a group of data. In other words, you can cut the number of things you have to do, make it easier to do these things, and cut the number of times you have to do them.

The type of equipment available to assist will frequently be an important consideration. Methods which seemed efficient and desirable for hand computations became obsolete and inefficient with the introduction of modern calculating machines. The graphic item counter of the IBM test scoring machine, various types of punched card, magnetic, and electronic procedures for recording and operating on numbers play an important role in most current computational procedures.

However, in many studies shortcut procedures which use no equipment except pencil and paper are very satisfactory, and actually faster and more efficient in some cases than procedures using the elaborate and awe-inspiring new types of machines.

To return to the problem of shortcuts for correlation computation procedures, the first thing suggested was to cut the number of operations to be performed. If, in some way, the data can be gotten into standard score form with means of zero and standard deviations of one, the computations involve only the summing of the multiplied pairs of values and dividing by their number. The second principle is to make it easier to do these operations. This suggests using simple numbers such as 1, 2, $\frac{1}{2}$, and 0. The third principle suggests finding simpler operations to perform. Counting the number of cases falling above or below some fixed value and comparing these numbers with tabled values of those that would be found in correlation surfaces having known coefficients appears much simpler than multiplying, summing, and dividing. The fourth principle involves cutting the number of times each operation is to be performed. This suggests grouping. Two groups might be used but it is found more efficient to use three, five, seven, or some odd number so that the middle group has a mean standard score value of zero.

These considerations have led to the development of two basic types of techniques. One involves the use of coarsely grouped standard scores so that the numbers multiplied are simple, such as 2, 1, $\frac{1}{2}$, 0, $-\frac{1}{2}$, -1, and -2 (**5, 11**). The other type of technique involves counting the cases in a specified tail of one of the distributions which exceed a specified value on the other distribution (**2, 4, 5**). This value and the corresponding value at the other end of the distribution are used to enter a two-way table to find the value of the correlation coefficient in a surface which showed these proportions in this part of the area.

Some years ago Kelley (**8**) showed that if upper and lower groups are to be used in which the cases are to be counted, the optimal size groups for discriminating between cases at the two ends of the distribution are 27 per cent. The present author prepared tables for obtaining correlations for such groups in 1936. A chart for the same purpose was published in 1939 (**3**).

One of the disadvantages of the procedure is that the extreme cases at the two ends of the distribution are grouped in with others closer to the mean and near the first and third quartile points respectively and do not receive the greater weight they deserve. To remedy this the present author has recommended (**1**) that the values nearer the extremes be counted twice to give them double weight in determining the appropriate proportions.

In the fall of 1950 it seemed that it would be desirable to get an accurate determination of the optimal proportions to be used if the extreme categories were each to receive double weight as compared with the cases in the next category. The most suitable procedure appeared to be to obtain those five categories which would maximize the proportion of the variance in the between groups variance as compared with the within group variance. By a series of successive approximations it was found that this was achieved by using groups of 9, 20, 42, 20, and 9 per cent. The cases in the extreme groups of 9 per cent are each counted as two in adding them to those in the next groups of 20. The middle group of 42 per cent is ignored.

Tables have been prepared giving the correlation coefficients which would have various proportions in the upper and lower 29 per cent of the distribution when the extreme 9 per cent is given double weight.

Similarly, it has been determined that if the most extreme cases are to be given triple weight and the next double, as compared to the larger groups, the maximum between groups variance will be obtained by using groups of 4, 8, 25, 26, 25, 8, and 4 per cent respectively.

It is of considerable interest to note just what proportion of the variance is between groups variance in each of the three cases described above. For upper-lower 27 per cent the value is 81.0 per cent. This value was found to be the maximum for equally weighted upper-lower groups, confirming Kelley's previous finding. For upper-lower 9, 20, the value is 90.9 per cent. For upper-lower 4, 8, 25, the value is 94.6 per cent. Thus it appears that only a small portion of the available information is lost by this type of grouping. It is interesting to note that the double weighting of the top 9 per cent regains more than half of the variance lost by using only the single upper and lower 27 per cent groups.

Tables have been prepared by the author for each of the above conditions.

When the data are on answer sheets and the test scoring machine can be used, these procedures are very efficient. In some instances where the data are in different form the use of coarsely grouped standard scores is desirable. Using the Kelley Statistical Tables, it is found that the upper 6 per cent of a normal distribution has a mean deviation of 1.985. The next 21 per cent has a mean deviation of 1.008. The next 8 per cent has a mean deviation of 0.497. The middle 30 per cent has a mean deviation which, of course, is zero. These coarse groups have a between groups variance of 93.9 per cent of the total variance. In the spring of 1950 tables were prepared for using this method to compute biserial correlation coefficients in connection with two projects with which the present author was connected. One of these (11) was a study for the United States Civil Service Commission on which the author was consultant to personnel of the Cincinnati Regional Office. The other (9) was a report prepared for the Personnel Research Section of the Adjutant General's Office, Department of the Army, by the department of psychology, University of Pittsburgh. The table used in these studies was copyrighted by the author and published by the American Institute for Research in September 1950 (5). A graph which can be used for the same purpose was prepared by personnel of the Test Development Section, United States Civil Service Commission, and published by *Psychometrika* in June 1951 (6).

At a cost of only a small portion of the available information (4 per cent of the total variance) the two 8 per cent groups may be added to the zero group. This group will contain 46 per cent of the cases and the computations will be appreciably simplified because of the increase in the number of zero crossproducts and the elimination of fractions.

EFFICIENCY OF SHORT METHODS

Data are available regarding the time required by these procedures. For product-moment coefficients using coarsely grouped data in these seven categories Zagorski (12) and Hahn (7) computed four or five hundred coefficients based on about 100 cases each. The time required averaged about five minutes per coefficient including all operations. It is of special interest to note that these operations were all performed using only pencil and paper and without the aid of mechanical or electrical computing machines.

Under Lange's supervision college students computed 8,000 biserial correlation coefficients using the coarsely grouped normalized standard scores for samples of two to three hundred cases (**9**). These coefficients required about two minutes each including all necessary computations.

Several thousand coefficients have been computed under the author's supervision using his upper-lower groups tables. These require less than one minute each including all operations with the counting done on the graphic item counter of a test scoring machine.

It seems safe to conclude that these times represent costs substantially below those usually incurred in calculating correlation coefficients. Under some circumstances the form of the data or other conditions may make other procedures more efficient. However, for a wide variety of situations these shortcut methods seem to offer substantial savings.

ACCURACY OF SHORT METHODS

The question remaining is: "How accurate are they?" Data bearing on this point were collected by means of an empirical sampling study on which the author was assisted in the early phases by Miss Leone Chesire.

Populations of 10,000 cases each in which the correlations were .00, .45, and .90 were set up with the help of Pearson's tables (**10**). Each case was given a number and 100 samples of 100 cases each for the three populations were drawn by using tables of random numbers including Tippett's tables and a set of tables kindly supplied by Dr. Joseph Berkson.

The coefficients were then obtained using a number of computational procedures. The results are shown in the Table 1. It is seen that the means of the 100 values are in all cases fairly close to the population values. On the whole, the tetrachoric coefficients show the greatest discrepancies. The most crucial values to be observed in the table are those in the three columns headed σ_a. These represent the actual standard deviations of the distributions of 100 coefficients as computed for each of the populations by the various methods. The tetrachoric coefficients are the least accurate in all cases. For the population in which the correlation is .90, the method of upper-lower 16 per cent shows little variability when there are only 100 cases, because in so many instances all of the cases in the upper 16 per cent are in the success group and vice versa for the lower 16 per cent.

The standard error for the upper-lower 27 per cent procedure compares very favorably with biserial r. Biserial r obtained from normalized coarsely grouped scores actually is smaller than the biserial r calculated by the more conventional procedures. The differences are very slight and are likely to represent the operation of chance factors which would

TABLE 1

EMPIRICALLY OBTAINED SAMPLING ERRORS OF CORRELATION COEFFICIENTS COMPUTED BY VARIOUS METHODS

Computational Procedure	*100 Samples of 100 Cases Each From A Population in Which* r = .00				*100 Samples of 100 Cases Each From A Population in Which* r = .45				*100 Samples of 100 Cases Each From A Population in Which* r = .90			
	M_r	σ_a	σ_t	SE_σ	M_r	σ_a	σ_t	SE_σ	M_r	σ_a	σ_t	SE_σ
Product-Moment	−.006	.086	.101	.007	.455	.078	.081	.006	.897	.020	.019	.001
Biserial *r*	−.007	.108	.125	.009	.444	.117	.106	.008	.903	.043	.044	.003
Biserial *r* (Normalized Groups)	−.008	.105	.125	.009	.449	.115	.104	.007	.911	.040	.043	.003
$r_{U4.8.25L4.8.25}$	−.008	.111	——	.008	.450	.109	——	.008	.896	.037	——	.003
$r_{U9.20L9.20}$	−.012	.113	——	.008	.448	.115	——	.008	.893	.036	——	.003
$r_{U_{27}L_{27}}$	−.005	.131	——	.009	.452	.119	——	.008	.892	.037	——	.003
$r_{U_{16}L_{16}}$	.002	.123	——	.009	.461	.132	——	.009	.880	.021	——	.002
$r_{\text{tetr.}}$	.021	.186	.157	.013	.436	.149	.134	.010	.889	.055	.048	.004

M_r = mean value of 100 coefficients.
σ_a = standard deviation of 100 coefficients.
σ_t = standard error of coefficient from formula.
SE_σ = standard error of the standard deviation for these samples.

be unlikely to be reproduced in another set of samples. For those coefficients for which it has been possible to derive sampling error formulas, the observed values agree fairly well with the theoretical values.

In conclusion it may be said that these data suggest that in many situations it would be desirable (*a*) to plan the study so as to include a few more cases than would be necessary to obtain the desired degree of sampling stability, and (*b*) to take advantage of the savings available in these short methods with no loss in sampling accuracy in those cases where a small number of additional cases can be secured to offset **the** slightly greater precision of the longer computational methods.

REFERENCES

1. BARHAM, DOROTHY. Incomplete sentences as a group projective technique. Unpublished master's thesis, Univer. of Pittsburgh, 1948.
2. FLANAGAN, J. C. *Factor analysis in the study of personality*. Palo Alto: Stanford Univer. Press, 1935.
3. FLANAGAN, J. C. General considerations in the selection of test items and a short method of estimating the

product-moment coefficient from data at the tails of the distribution. *J. educ. Psychol.*, 1939, **30**, 674–680.
4. Flanagan, J. C. Item analysis by test scoring machine Graphic Item Counter. In *Proceedings of the Educational Research Forum*. Endicott: 1940.
5. Flanagan, J. C. *A table for obtaining the biserial correlation coefficient.* Pittsburgh: American Institute for Research, 1950.
6. Goheen, H. W., & Davidoff, M. D. A graphical method for the rapid calculation of biserial and point biserial correlation in test research. *Psychometrika*, 1951, **16**, 239–242.
7. Hahn, C. P. Predicting job success of selected department store employees. Unpublished master's thesis, Univer. of Pittsburgh, 1951.
8. Kelley, T. L. The selection of upper and lower groups for the validation of test items. *J. educ. Psychol.*, 1939, **30**, 17–24.
9. Lange, C. J. Evaluation of procedures used to develop personality and interest items. Department of the Army, Personnel Research Section Report No. 876, Department of Psychology, Univer. of Pittsburgh, 1951.
10. Pearson, K. (Ed.) *Tables for statisticians and biometricians*, Part II. London: Biometric Laboratory, Univer. College, 1931.
11. Wager, C. E., & Sharon, M. I. Revision of Test for Machinist based on item analysis of Machinists Test #1607, Series I. In *Quality Examining Program* of the Sixth U. S. Civil Service Region. Cincinnati: June 1950.
12. Zagorski, H. J. A factor study of self-evaluation. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1951.

Received September 18, 1951.

NOTICE

Dr. Wayne Dennis will become editor of the *Psychological Bulletin* as of January 1, 1953. All new manuscripts should *henceforth* be sent to him at the following address:

PSYCHOLOGICAL BULLETIN
Dr. Wayne Dennis, Editor
Department of Psychology
Brooklyn College
Brooklyn, New York

The Book Review Editor will be Dr. Edward Girden. All books and other publications intended for listing or review should be addressed to him as follows:

PSYCHOLOGICAL BULLETIN
Dr. Edward Girden, Book Review Editor
Department of Psychology
Brooklyn College
Brooklyn, New York

BOOK REVIEWS

HELSON, H. (Ed.) *Theoretical foundations of psychology*. New York: Van Nostrand, 1951. Pp. xix+787. $7.25.

For many, this book will have an exciting title. *Theoretical Foundations of Psychology* could be a book which examines the methodological presuppositions of psychological theory in the sense that inquiries into the "foundations" of mathematics perform a similar function for that discipline. Such a book is certainly needed. Or, the title might imply an assessment of emerging common properties within the untidy matrix of current theoretical efforts. All would welcome an attempt of this sort. The title could suggest a general comparative analysis of current theories. A craftsmanlike contribution of this order would be useful. Finally, the title might mark the intention to formulate, develop, or evaluate the "fundamentals" of some *particular* theory. Such an effort might at the least prove interesting. Actually, this lengthy symposium is all of these things and none of them; it is, in addition, occasionally a primer in general psychology, at times a supplementary survey of "fields of psychology," often a *Psychological Bulletin* review article, and, in one or two instances, a *salon* in which a contributor hangs a one-man show of his *Lebenswerke*.

Professor Helson does not quite succeed in preparing the reader for all of this in his *Preface*. We are informed that this "book is designed to treat the fundamental ideas, concepts, theories and problems which are at the center of the chief divisions of the subject." Further, the "volume should help advanced undergraduate and graduate students . . . to achieve some measure of integration of particular fields as well as psychology as a whole." Even the "mature psychologist, familiar with the best thinking in his own field will welcome . . . basic treatments of other fields." These aims were apparently realized by "the splendid cooperation of the contributors who . . . have succeeded in producing a volume fairly homogeneous as to level and type of material." To this reviewer, the only "homogeneity" evident in the volume was precisely in the single sense in which *heterogeneity* was promised. Despite Professor Guilford's comment in his brief *Foreword* (as general editor of the Van Nostrand series) that "we have passed beyond the school-forming era and that lines of affiliation are no longer clearly drawn," the one uniformity common to most of the seventeen contributors seems to be a predilection for some kind of theoretical framework of the sort variously called "field," or "neo-gestalt," or "dynamic," or "holistic."

It may prove instructive to approach this large, uneven volume as a kind of clinical exhibit of certain general trends in present "theoretical" psychology, and in psychological writing generally. As in any clinical exhibit, such trends may be expected to emerge in exaggerated and prototypical form.

One trend which the book defines with disturbing impact is the superficial level at which vital methodological issues and topics from the philosophy of science are discussed. Those who have been a little concerned about the superficiality of much of the recent methodological literature can only find their anxieties confirmed within this volume. The tone is set in the first chapter, *Methodological Considerations*, by M. G. Preston. This chapter consists of an arbitrary stringing together of assertions about methodological and theoretical issues, half of them erroneous or distorted, and half deriving from the cliché-box of contemporary methodology. The tenor of the discussion may be inferred from the following sample of how major methodological concepts are dispatched: "When an assertion is made on the basis of faith it is called a postulate" (p. 7). Among the high lights of the presentation are a treatment of operational definition in which the operational criterion is nowhere specified, an obscurely formulated distinction between "reductionists" and "non-reductionists" which apparently represents the author's principal schema for the classification of psychological theorists, and a curious discrimination between "theories which presuppose inductive processes and theories which presuppose deductive processes" (p. 19). A treatment of independent, intervening, and dependent variables fails to distinguish between the senses in which "independent" and "dependent variables" designate data language specifications of the variables which are manipulated and recorded in concrete experiments (*experimental sense*), and theory language identifications of the chief empirical variable-classes which the theory is designed to relate (*theoretical sense*). This leads to the development of the somewhat labored theses that "Some dependent variables are called intervening variables" (pp. 16–18), and "intervening variables may be used as independent variables" (p. 19).

The first chapter presages the level of methodological analysis maintained in most of the others. Each author, however, adds his individual stamp in the rendering and evaluation of central methodological concepts. Terms like "molar," "molecular," "field," "intervening variable," "operational definition," "theory," etc., become chameleons which assume new semantic and emotive shadings in each chapter. Take the case of "operational definition." At one end of the spectrum is S. H. Bartley who, at the end of a chapter (*Fatigue and Efficiency*) devoted largely to a demonstration of the evils of regarding fatigue "as a narrow phenomenon to be defined 'operationally'" (p. 336), finally finds the *real* fatigue in the "layman's" knowledge of his "own feelings." At the other end of the spectrum lies J. Loevinger's chapter on *Intelligence* which, after dismissing as blind alleys the "philosophical," "pragmatic," and "factorial" solutions to the problem of intelligence, arrives at an "operational solution." Or take "intervening variables": Krech in his chapter, *Psychological Theory and Social Psychology*, apparently feels that a wondrous reorientation of social psychology ensues immediately upon the remarkable intuition that "beliefs and attitudes" are "intervening variables"; for Beebe-Center (*Feeling and Emotion*) the "intervening variable" functions, in his extended analysis of the "construct" of "palatability," as a laboriously developed shorthand device for summarizing interrelations within a restricted domain of experimental data.

Attitudes towards "theory," in this remarkably comprehensive book, range all the way from Krech's, and MacKinnon and Maslow's (*Personality*) espousal of the global, if "programmatic" framework, to Harlow's (*Thinking*) comfortable basking in empiricism. Between these extremes, we find Helson (*Perception*) advocating a type of theory, instanced by the theory of "adaptation-level," which, though specified in scope, envisages the large aims of "supplementing," "broadening," and quantifying previous theories; we find also Beebe-Center, who seems on the side of the narrow scope systematization of highly restricted experimental domains.

More distressing than the range of the authors' *attitudes* towards theory, or their explicit *views* on problems of theory construction, are the occasional embodiments of these attitudes and views in concrete theoretical suggestions. For instance, E. B. Newman (*Learning*), after rejecting a series of one-paragraph résumés of contemporary theories of learning, offers to "make up the deficit [he] may have created" (p. 436), in a final section called *Suggestions for a Theory*. He does this in terms of six "theoretical considerations," each of which (p. 437) "might be called a principle, but actually . . . is more like a cluster of principles." Two of these "considerations"—"sensory selectivity" and "behavioral parsimony"—seem to mean what they ordinarily mean when used by other writers in the "field" tradition. Three others—"carryover," "coalescence" and "response patterning"—seem to be words which, for some reason, are preferred to equivalent terms employed by field theorists, such as "traces," "trace organization," and "insight." The unique contribution is "motivational spread" which "represents an attempt to give a specific interpretation of . . . reinforcement" (p. 437), but the meaning of which requires further interpretation. How these six "principles," or "cluster(s) of principles" or "headings" or words are to be interrelated within the theory the author has in mind, what their systematic status is or is expected to become, are questions outside the scope of Newman's discussion. A comparably inchoate conception of theory seems implicit in Bartley's specification of a programmatic "theoretical structure" for the investigation of fatigue, and Krech's adumbration of a "'perceptual' theory" for social psychology. If the reader of this volume is likely to receive less than enlightenment by the *talk about* theory, his situation will not be improved by the "theorizing" actually attempted.

Another aspect of the book which seems disturbingly symptomatic of a general shortcoming, not only in psychological writing and pedagogy, but in thinking as well, is the inadequate treatment or even total neglect of problems having to do with the interrelationships between each field and the others. The "advanced undergraduate and graduate students" who are trying to get some kind of view of "psychology as a whole" will, instead of achieving the "measure of integration" promised them, find only that their initial obfuscation is confirmed and amplified. Failure to consider problems of interrelationships in a sustained and rigorous way has consequences which transcend the merely pedagogical; any approach towards adequate theory demands, at every stage, tough-minded attempts to block out the possible ways in which discrete research areas mesh (or do not mesh) with each other and with the objectives of the science.

A book called *Theoretical Foundations of Psychology* should place a particular emphasis on meeting this obligation. Instead, we find at least four contributors—J. Oppenheimer (*Some Problems of Nervous Function*), E. Girden (*Some Neurological Correlates of Behavior*), N. Bayley (*Development and Maturation*), and J. Loevinger (*Intelligence*)—who make virtually no attempt to deal either with interrelations between their fields and others, or with the place of their specialties in the science as a whole. Of the remaining contributors, some (e.g., D. B. Klein on *Abnormal Psychology*, MacKinnon and Maslow on *Personality*, Harlow on *Thinking*) have made only a token effort to deal with interrelationship problems. Others (e.g., Newman on *Learning*, Helson on *Perception*) have touched on such problems in essentially stereotypical terms. A small minority—Krech and Beebe-Center—have subjected such questions to tortured, but somewhat empty, scrutiny.

While most of the authors slight interrelationship issues, each in his own characteristic way, it may be useful to discuss some concrete instances. Representative of the authors who neglect these issues almost entirely is Bayley, who proceeds to her survey of information about *Development and Maturation* after merely noting in a single sentence (p. 145) that "an understanding of psychological functions is greatly enhanced by knowledge about the processes through which these functions develop . . . with . . . age." Loevinger's excellent critical appraisal of factorial and other avenues to the analysis of intelligence could have gained in stature if, at some point, "intelligence" had been related to the problem of individual difference variables within more general theories of behavior, rather than approached as the circumscribed *Ding an sich* whose elusive ontological properties seem forever under pursuit by the investigators in this area. As representative of the contributors who provide only token analyses of interrelationship problems, we find MacKinnon and Maslow repeating endless variations on the theme that "investigations of personality . . . now make mandatory a revision of general psychological theory" (p. 602). At no point, however, do we find concrete hints as to what these revisions may be, nor indeed what concrete relations do or should obtain between general psychology and personality. Newman, the symposium's expert on learning, and Helson, in his chapter on perception, offer us little more with respect to interrelationship problems than an emphasis, congenial to most with gestalt-like sympathies, on the intimate relations between the phenomena and principles of perception and learning—perception, of course, having systematic and functional priority. Krech, one of the few contributors who makes the assessment of interrelations central, emerges from some forty pages of text with the conclusion that "General psychology is social psychology" (p. 664), a conclusion which seems to confer another dimension of ambiguity on the analysis of the verb "to be" as treated in formal logic. Although this reviewer would guess at Krech's chief difficulty only with timidity, it is difficult to avoid the feeling that his discovery that all theoretical variables of general psychology are saturated with "sociality" overlooks the perfectly legitimate distinction between general laws and the conditions under which they operate. Beebe-Center devotes the substance of his contribution on feeling and emotion to a serious attempt to create a place for his topic within psychology. His earnestly documented theme is that progress in the conceptualization of feeling and emotion "during the last quarter of a century appears to have been the substitution of construction for reification" (p. 295). Nevertheless, when Beebe-Center comes to the "earthy" modern behavioristic theories of his preference, he fails to tell us precisely which elements of the constructions are relevant to the phenomena denoted by the now-outmoded reifications.

A third feature of the book, not without symptomatic significance for psychology in general, has to do with a complex of factors epitomized by the nature and level of the writing, by organization, and by style. Those who occasionally face the unhappy task of helping graduate students approximate to clear, scientific prose have reason to be worried by the prevalence of models similar to the present volume. The net impression created by the majority of chapters is that of turgid, imprecise, often vacuous writing. The fact that the personal flavor of the prose varies from the pseudo-Jamesian "Harvard-psychologese" ("memories seem . . . linked . . . like kaleidoscopic sausages in a giant string" [p. 395]; "I have . . . stretched the immensely complex past on a Procrustean bed" [p. 405]) of the chapter on learning, through the shabby rhetoric of the chapter on social psychology, to the inflated verbosity of the chapter on abnormal psychology, is secondary. More serious is the frequent failure to meet minimal standards of scientific or scholarly writing. At its worst, this may involve the failure to communicate anything; at its best the failure to communicate anything worth while.

Perhaps the most common form of this failing is in the "level" at which experiments are summarized in the many "extant evidence" surveys and "relevant example" references which the authors feel compelled to include. Such summaries are nearly always far too skeletal, allusive, or vague to have meaning for those who are not already familiar with the study; nor do they serve better than the mere mention of the investigator's name to reactivate the memory traces of those who are familiar with the material. Of the chapters which include substantial reference to experimental materials, only the chapters on *Thinking* and *Some Problems of Nervous Function* manage to avoid this expendable level of summarizing. Equally, if not more pernicious, is the almost universal tendency to summarize complex theories in terms of an inexact translation of a single principle, or, at best, an impressionistic two-paragraph distillation of "essentials."

That these considerations of "style" are not irrelevant to content, can, I think, be proved by something like a random selection of quotations. One should suffice. The following passage appears somewhere in the book under the heading, "The Essential Unity of Psychology":

> Irrespective of the difference in their views, Lashley and Hull are in agreement on the fact that psychologists, like all scientists, are engaged in the work of studying the interrelations among concepts, in particular, their functional interrelationships. Secondly, they are in agreement that the work of studying such functional relationships suffers in the extent to which it fails to meet the requirements of the experimental method. All psychologists are united in the study of behavior by means of the investigation of the functional dependence of concepts, through the use of the experimental method.

The tone of the present review should lend sincerity to some final comments about the happier ingredients of the volume. The really exceptional chapter in this book is Guilford and Comrey's judicious, knowledgeable introduction to *Measurement*. Instead of plunging into a statement of the formal conditions of measurement as in most primer presentations, the approach is to exhibit the significance of such formal criteria by first showing the distinction between pure and applied mathematical systems, and then treating measurement as the problem of finding specific empirical interpretations of the system of rational numbers. A second outstanding chapter—*Some Problems of Nervous Function*—is contributed by the symposium's only nonpsychologist, J. Oppenheimer. This well-organized, literate introduction to some foundation concepts of neurophysiology, with its refreshing emphasis on up-to-date research and its freedom from the usual stereotypes of the introductory text, should prove a useful reference in connection with serious undergraduate courses.

A number of additional chapters, although not untainted by the general trends previously noted, make, in some way, useful contributions. Despite Loevinger's failure to consider the problem of intelligence within a more general theoretical context, the bulk of her chapter is devoted to a skillful teasing out of the hidden apriorism of factor analytic approaches to intelligence. Helson's discussion of perception has the merit of some incisive analysis of a number of historical and recent theortical issues. The outline of his theory of adaptation-level is, however, too schematic to be intelligible to any but those who, by virtue of previous knowledge, would find such a précis supererogatory. However slim Harlow's chapter may be in theoretical content, he does provide us with a useful, if selective, survey of research on the formation and functioning of "organized response patterns." It is regrettable that he nowhere tells us what "organized response patterns" means. F. W. Irwin's chapter on *Motivation* compares favorably with the chapters on this topic in most introductory texts, but is of uneven excellence. It suffers especially from a casual rendering of motivational "theories" and haphazard organization, but contains several thoughtful sections (e.g., the discussion of primary drives).

The word "theory" in contemporary psychology is a prestige word. It should be. In its best sense, it denotes a value which any science should be proud to work towards or attain. But in psychology and the social sciences this word is in danger of becoming a "prestige symbol" in the unhappy sense of the device used by advertisers to facilitate acceptance of a product, or by progagandists to produce feelings of awe. Detailed appraisal of a book so conspicuously unsatisfactory as the one under review may seem unnecessary or trivial. But at a time in psychology when nothing is more important than disciplined assessment of the preconditions to sound *theory*, it is fitting to question any further vulgarization of the term. It *could* just happen that someone might interpret *Theoretical Foundations of Psychology* as being relevant to theory.

SIGMUND KOCH.

Duke University.

MARX, MELVIN H. (Ed.) *Psychological theory*. New York: Macmillan, 1951. Pp. xi+585. $5.00.

This volume is of interest not only for what it contains, but for raising the question of how we can best introduce students to psychological theory. Professor Marx proposes accomplishing this by means of a course built around a collection of readings from works of contemporary theorists.

In brief, the book contains selections from journal articles and books, organized under plausible section headings and threaded together by means of brief editorial comments. Part I includes selections concerned with methods and principles of theory construction in psychology while Part II is made up of representative writings from contemporary theoretical works.

The first part is planned "to encourage a more intimate understanding and a sounder utilization of the principles of theory construction." These aims will certainly be at least partly realized. A number of well-known and influential papers—concerned, for example, with operationism, types of constructs used in psychology, levels of explanation, interbehavioral psychology, and applications of topology, factor analysis, operational analysis, and hypothetical-deductive methods in psychology—are made readily available to the student. The editor's introduction to this series of papers emphasizes the empirical basis of theory. He corrects the unhappy tendency of a number of psychological writers to identify operationism as a formal rule for defining acceptable concepts. Theories based upon infra-human research are defended vigorously. It is refreshing to see someone frankly take the stand that it may be respectable to study animal behavior simply with a view to understanding the behavior. This presentation may be better suited to engage the interest of the potential scientists in our classes than the common textbook justification of animal research as part of a broad program calculated to pay off in effective study habits, irresistable advertising methods, and popularity for everybody.

The notion of utilizing "principles of theory construction" is somewhat ambiguous. "Principles of criticism" would be more descriptive. It is safe to expect, however, that Part I of this volume, if it does not teach the reader how to construct theories, will at least influence him in the direction of being better able to appreciate good theories and less apt to publish bad ones. These readings will be good to have in the hands of students, especially since the editor has included the original lists of references with the reprinted papers, a device which may lure many readers into a more serious study of the method and philosophy of science.

The stated purpose of Part II is "to give the advanced student a certain amount of first-hand experience with the work of the various men who have laid the theoretical foundations for contemporary psychology within the past two decades." The twenty-four selections are drawn from the areas of perception, learning, psychodynamics (psychoanalysis, psychosomatics, psychological study of conflict and frustration), personality, and social interaction. The second part does not seem to be on a par either quantitatively or qualitatively with the first. The twenty-three selections in Part I average approximately sixteen pages in length; they are nearly all relatively self-contained papers that could readily form the bases for group discussion in classes or seminars. The twenty-four selections in Part II average only seven and one-half pages in length and are for the most part excerpts from larger works. These readings are said to be representative of the theorists in question. To illustrate the scale of sampling involved we note that the work of Koffka is represented by a nine-page excerpt from *Principles of Gestalt Psychology*, that of Freud by five pages from *The Problem of Anxiety*, that of Skinner by approximately one-third of his early paper on the concept of the reflex. The array of theorists chosen for inclusion appears to reflect conventional standards quite accurately, although there are a few striking omissions. Some psychologists may not be happy with the impression that all theory of perception is gestalt or that there are no influential theories in the entire area of physiological psychology. On the other hand, it is not clear what could have been done to fill the gaps. The volume is large as it is, and the readings included could hardly have been cut much further without losing both syntax and sense.

The editor has deftly cut the ground from under potential critics of his selections by indicating in a Preface that they reflect his judgment neither with respect to "ultimate scientific significance" of the theorists nor to their value in exemplifying the view of theory construction presented in the papers of Part I. The product of this rather aloof editorial policy is an anthology that may be better fare for browsers among mature psychologists than for students receiving their first introduction to theorists and theories. It is doubtful that either theorists or theories can be usefully presented for instructional purposes by means of a strictly cross-sectional anthology of this sort. A scientific theory is an on-going activity, and its essential features are revealed only as it is seen in the process of development, modification, and interplay with experiment.

The possibility of developing the student's critical behavior is not strengthened by the plan of Part II. For on what grounds is the student to accept or reject the principles and points of view concerning theory construction presented in the first part? Presumably his judgment should be based upon observation of the consequences in actual practice when these principles are used or misused. But for this, we had as well skip Part II and send him to the library.

In general outline, this book reflects to some extent an accelerating shift in interest among psychologists from system building on a grand scale toward the development of more technical, experimentally oriented theories. The instructor who wishes to keep pace with this trend by developing new techniques for teaching psychological theory will find that Professor Marx has contributed a welcome experimental approach to the problem.

WILLIAM K. ESTES.

Indiana University.

MUNN, NORMAN L. *Psychology: The fundamentals of human adjustment.* Pp. vxi+624. $4.50. 2nd edition. Boston: Houghton Mifflin, 1951.

The second edition of this now famous text is so nearly the same as the first (see Helson's review, this JOURNAL, 1947) that one wonders why Professor Munn felt it necessary to revise the book so soon. As an eclectic but scientifically sound work it has few competitors for use in one-semester courses of the general survey type.

To instructors attempting to introduce students to theory, however, the book's lack of theoretical discrimination is bound to be unsatisfying if not down-right irritating. Eclecticism in textbooks is essentially messy: neither the student nor the instructor ever quite gets closure on anything. Helson made a similar criticism on the first edition when he said it lacked an underlying point of view. This defect has not been corrected. Instead, Munn has added here and there some journalistic reporting of certain current theoretical issues. This is a good thing to try to do, if done well. Unfortunately, Munn doesn't do it well.

His big attempt is in the field of learning where he devotes eight pages to comparing the theories of Tolman and Hull. Reading these, the student will become confused rather than excited and the instructor will suspect that Munn doesn't understand the purpose of theory. He starts out with what seems to be a clear differentiation between the constructs of Hull and Tolman, but the distinction is lost as it becomes clear (to Munn) that Hull, after all, has only responses—and old ones at that—whereas Tolman has meanings and knowledge and cognitive maps. Gone are Hull's intervening variables in a wave of sentences like "One additional difficulty with the idea that *all learning is reducible to conditioned responses* is that, in order that a response may be conditioned, it must already be present" (p. 189, italics mine). A pseudo-resolution of the theoretical conflict is offered: both responses and knowledge are acquired in learning! One wishes Professor Munn had, after all, stuck to facts and never brought the matter up.

Incidentally, Munn has a new definition of psychology: it is the science of behavior, period. In the first book we studied "behavior and experience." We still do, really. That is, Munn is still quite as eclectic about subject matter as ever and by no means can the book be called behavioristic. Neither can it be called mentalistic, structuralist, gestalt, phenomenological or anything else, except possibly physiological.

So much for theory. The organization of the book is unchanged except for numerous new section headings, expanded introductions to the Parts and some condensation. The book now numbers 624 pages instead of 497 but this is an illusory increase in size; actually the type is bigger (and clearer) and most of the earlier fine-print sections have been incorporated into the regular-print running text—an admirable improvement. There is an absolute increase of 76 in the number of illustrations, 302 as against 226 in the first edition. Most of these are wonderful and many are new. In fact, the illustrations are the very best features of the book and help enormously to bring psychology alive.

A sample of other changes is the following: the facial vision studies now take precedence over the pipe-smoking experiment as the prototype of experimental method. A timely section on human engineering has been added. Unfortunately, the sections on neural summation and inhibition have been omitted and the section on retinal interaction abridged to meaninglessness, leaving, actually, no reference when page 435 (retinal interaction) says to refer to page 47 (neural summation) because page 47 no longer has anything to say on the matter. There is a new treatment of film and aperture modes of color but perhaps not enough to satisfy Helson. World War II data now replace old Army Alpha where necessary. The Wechsler-Bellevue is described and the measurement of infant IQ is properly added. Emotion as a conscious experience is deleted, in line with the "new" definition of psychology, I suppose. The new combination chapter on *Acquired Motives* is a competent rewrite job, but says little about theory of acquisition and neglects entirely the work of Miller and others, especially on fear and anxiety. But then Lewin and Maier are reduced almost to footnote status. Without counting words, I estimate the changes in the motivation part to be less than five per cent and this change probably is typical of the book as a whole.

One proper conclusion: if your bookstore is overstocked in the first edition, your can still feel safe in using it. Nevertheless, the new edition is a slight, though probably statistically insignificant improvement over the first edition.

In the face of valid criticism, why should anyone want to use the book as an introductory text? Because it is a fair picture of modern psychology. It is just about as sound and as confused, as brilliant and as dull, as heterogeneous and disorganized as the science of psychology itself is, in this year of 1952. It doesn't *bring* anything to our science; it just photographs it. The photo is not tinted with the charm of a William James; it is not as systematically narrow as a Watson or a Titchener; and it isn't as empty as Woodworth. It is a beautifully prepared and attractive document. Students like and respect it and most of what they learn from it will become a sound foundation for advanced study.

STANLEY B. WILLIAMS.

College of William and Mary.

STONE, CALVIN P. (Ed.) *Comparative psychology.* (3rd Ed.) New York: Prentice-Hall, 1951. Pp. xvii+525. $6.00.

If Frank Beach is right in holding that comparative psychology in this country has vanished from having concentrated exclusively on the white rat (Snark, species Boojum), then this book goes only a little way towards resurrection of the departed. It is the third edition of the well-known *Comparative Psychology* previously edited by F. A. Moss. Contributing authors to the new edition are Nicholas E. Collias, Calvin S. Hall, Harry F. Harlow, William T. Heron, Donald G. Marquis, Henry W. Nissen, Robert A. Patton, Karl U. Smith, Kenneth W. Spence, Calvin P. Stone, Rolland H. Waters, Paul T. Young, and William C. Young—an imposing array of authors!

Some fine new sections on social and abnormal animal behavior have been added and the other chapters brought up to date, but the general format of the book remains unchanged. Individually written sections deal with such classical areas of psychology as maturation, motivation, learning theory, and discrimination. This should cause the true comparative psychologist immediately to suspect the snark: that despite the qualifying term "animal" before each section heading, here hides a general psychology text.

What is the difference? Well, unless we are all to be regarded as comparative psychologists (everyone works with some kind of animal), the term "comparative"—apparently a somewhat honorific term—should be reserved for those who *compare* the behavior of diverse animals, that is, do studies which in the words of Editor Stone "chiefly . . . are intended to bring out differences and similarities." The diversity of animals given coverage in this book is limited almost completely to mammals. And of these, the inevitable white rat and a few species of primates crowd all others off the stage. How much comparison is done within this small range of animals? Relatively little. For evidence, count the number of figures in the book which graph for us behavioral data of any kind. There are thirty-five. Now out of these, count the number which present simultaneously on the same set of axes data to be compared for more than one kind of animal. There is just one. Maybe this is too rigorous a test, yet the climate of the book is such that at least one section could be included (Spence's) that gives not the slightest suggestion that the subject matter ("Theoretical Interpretations of Learning") would possibly have phyletic parameters or provisos. Most of the sections could be transplanted unaltered into an advanced general psychology text without causing eyebrows to be lifted. This stems from the fact that the authors are generally topic oriented rather than animal oriented.

Aside from any possible mistitling of the book (and this after all may be unimportant), a striking feature of the work is its heterogeneity in style and level of difficulty. Flesch scores must range from the very low for sections by Hall and P. T. Young to very high for the sections by Marquis and Smith. The various authors were obviously aiming at different-level students. For example, Hall assumes that his readers need to be introduced to the general notions of reliability and validity of tests, whereas Smith assumes a reader knowledge of the distinctions between the psychophysical method of constant stimuli and method of limits. Heron explans that Pavlov worked with dogs and spells out the meaning of conditioned and unconditioned stimuli, while Harlow and Spence take it for granted that the reader knows what a Skinner box is. Nowhere in the whole book is the Skinner box (a natural habitat for *Mus norvegicus albinus*) described.

There are still things to be done with a heterogeneous book. The reviewer has found occasion to assign portions of it to his undergraduate students in experimental psychology, and other portions to graduate students in a learning seminar. This may reflect the secret of the book's continued success. Although it is not likely to be used as a text in a genuine course in comparative psychology, it is useful as a reference text in a lot of courses. The authors are all eminent writers who have something to say, and we want our students to read them even if the book that contains them is not "of a piece."

DAVID ZEAMAN.

University of Connecticut.

MILLER, GEORGE A. *Language and communication.* New York: McGraw-Hill, 1951. Pp. xiii+298. $5.00.

The past decade has witnessed a rapid development of interest in communication problems, ranging from mathematical and engineering studies to large-scale studies on the effects of mass communications. For their own part, psychologists have become increasingly concerned with various aspects of language behavior—the nature of symbolic processes, intelligibility of messages, structure of verbal habits, and communication channels in groups of various kinds, to give only a partial listing. Courses on language behavior are beginning to appear on university curricula at both graduate and undergraduate levels. Yet we are severely handicapped by the lack of adequate texts and reference books in this area. The available books on language behavior are generally of ancient vintage and written from a frame of reference quite foreign to most contemporary students.

Miller's *Language and Communication* is a significant contribution as the first modern text in this awakening field. It is written by one of the most competent researchers on communication problems—Miller can speak with the authority of first-hand observation on many topics in this area. He makes comprehensible to the average student many implications of the mathematical communication theory developed by Shannon, Wiener, and others. As a source of information on what has been done and is being done by psychologists in studying language behavior, this will be an extremely useful book; there is no other compilation available with its breadth of coverage. This reviewer was also happy to find, early in the book, a brief but competent survey of linguistics; most psychologists have been oblivious to the contributions of linguistic analysis.

Miller lays down the general plan for his book in the introductory chapter: it is to be a *scientific* book on language, emphasizing quantitative data, and it is to be based on the mathematical model provided by Shannon's *information theory*. This blueprint is hewn to pretty closely throughout, but the emphases which develop at several points suggest that this model may not be entirely satisfactory. It is a very stimulating approach, however, and one which encourages rigorous thinking. The first part of the book (Chs. 1–5) provides the student with basic tools for the study of language behavior, including excellent treatments of information theory, physiological and acoustical phonetics, speech perception, and statistical techniques of language analysis. The second part (Chs. 6–12) covers more "psychological" aspects of communication, including individual differences, development in the child, the structure of verbal habits (associations), the nature of symbolic processes, and social communication.

On the negative side of the ledger, this reviewer feels that *Language and Communication* is not a well-balanced presentation. Its quality is excellent in those areas where Miller himself is expert but sometimes rather thin in other areas; similarly, in quantity it is extremely detailed on topics where Miller has specialized but often skimpy elsewhere. For example, many pages are devoted to masking and other acoustical distortions despite Miller's own conclusion (p. 69) that "not until nearly all of the clues are eliminated does the listener falter"; yet the broadly significant topic of abstraction and concept formation receives only a scant two pages. The psycholinguistic problem of *decoding* (reception and interpretation of messages) is treated excellently and in detail, but the complementary psycholinguistic problem of *encoding* (selection and production of correctly coded messages) is hardly touched upon. This imbalance results in a highly variable level of difficulty for the reader. I imagine even most advanced graduate students will find the condensed materials on hearing very heavy going—there is insufficient redundancy!—whereas most of the materials on verbal habits and symbolic processes could be readily handled by the average undergraduate.

A serious weakness in this book, from the viewpoint of this reviewer, lies in Miller's handling of the role of learning (particularly in Ch. 8). He depends entirely upon the simple, one-stage S-R model developed by Skinner. Since this model takes no account of intervening, mediating variables, it is patently incapable of handling most human language phenomena. The picture one gets of the human communicator is one of a cluster of sheer labeling associations, external stimuli calling forth vocal noises with nothing in between. This reflects Miller's understandable attempt to be rigorously scientific and to deal only with direct observables, but I feel that it is possible to be equally rigorous with a more elaborate and flexible learning theory model that encompasses mediation processes. As a matter of fact, the limitations in the Skinnerian model become apparent from Miller's failure to use it systematically in later chapters on verbal learning and symbolic processes. Another outgrowth of the limited learning model is Miller's treatment of *meaning*—or better, lack of treatment. The entire problem of linguistic forms as *signs* having representational properties, of the development and measurement of significance, is missing. This also presumably reflects Miller's belief that we cannot be scientific and objective about meaning.

In spite of these shortcomings, Miller offers a highly competent and cleanly written book on language behavior, whose objectivity is refreshing after plowing through the hazy murk of armchair theorizing that has characterized much of the writing in this field. As a textbook, it is factually rather than theoretically oriented and brings the reader into contact with a wealth of new and important material; as an exposition of a point of view, it illustrates how powerful an analytic tool the model developed by Shannon can be. I have included this text in a graduate course on the psychology of the communication process and it has proven to be a very valuable and stimulating addition.

C. E. OSGOOD.

University of Illinois.

LUNEBURG, RUDOLF K. *Mathematical analysis of binocular vision.* Princeton: Published for the Dartmouth Eye Institute by Princeton Univer. Press, 1947. Pp. vi+104. $2.50.

There are some advantages in reviewing a book five years after its appearance. By the end of that interval the reviewer can assess the initial effects of the work and take advantage of later expositions[1,2] to provide a widened context for his evaluation.

Luneburg's book promises to have a great effect on theories of space perception. It has evoked considerable interest and promises to contribute significant findings through research that it has stimulated. Luneburg's premature death is a great loss to programs generated by his writings and to science in general.

The present review attempts to do two things: it first presents a summary of Luneburg's theory (together with an account of some experimental methods useful for evaluating parameters) and second, it analyzes some implications of the theory for a systematic study of behavior. Because we feel that it is desirable to present a rather full account, our summary is given in more than the usual detail. We use the slightly modified symbolism of the conclusion of Luneburg's book and of his later papers.[1,2]

Luneburg's task in *Mathematical Analysis of Binocular Vision* is to determine the mapping of physical space into visual space and then to characterize the visual space. Both these tasks can be fulfilled if the metric of visual space can be expressed in physical coordinates.

The concept *visual space* requires elucidation. The visual space is a hypothetical space; Luneburg calls it the space of visual sensations. The visual space is not a physical space; it is not the space in which an observer manipulates experimental objects. When an observer says

---

[1] LUNEBURG, R. K. Metric methods in binocular visual perception. *Studies and Essays.* Courant Anniversary Volume. New York: Interscience Publishers, Inc., 1948.

[2] LUNEBURG, R. K. The metric of binocular visual space. *J. opt. Soc. Amer.*, 1950, 40, 627–642.

that one pair of points *looks* closer together than another, then in the visual space the first pair *is* closer together. The visual space differs from the continua usually obtained in psychological scaling procedures only in that it is three-dimensional.

If the mapping is determined, then an experimenter will be enabled to predict, from the physical coordinates of experimental objects, how an observer will say they appear with regard to form and localization, or, if the method of adjustment can be used, how the observer will set the experimental objects, i.e., their physical coordinates can be predicted.

A metric or distance function of a set of points is a non-negative number D satisfying three conditions. For any points, P_i, of the set: (1) $D(P_1, P_2)=D(P_2, P_1)$; (2) $D(P_1, P_2)=0$ if and only if $P_1=P_2$; and (3) $D(P_1, P_2)+D(P_2, P_3)\geqq D(P_1, P_3)$.

The metric of any space is given in differential form by Riemann's generalized line element. For the special case of a three-dimensional space with constant curvature K, this is $ds=[(dx^2+dy^2+dz^2)^{\frac{1}{2}}]/[1+\frac{1}{4}K(x^2+y^2+z^2)]$. Since the metric of a space determines all the properties of the space, the discovery of the metric of visual space will lead to its geometry.

Spaces represented by this differential, i.e., spaces with constant curvature, are the only spaces in which form and localization are uncorrelated, in which objects may be moved without suffering distortion in shape. Free movability, it may reasonably be assumed, is a characteristic of visual space. This conclusion restricts the nature of the visual space tremendously. If its curvature is constant, then Riemann's line element represents its geometry and there remain three possibilities: (1) $K=0$ (Euclidean geometry); (2) $K>0$ (elliptic geometry); or (3) $K<0$ (hyperbolic geometry). Which curvature adequately represents visual space can be answered only experimentally.

In addition to satisfying the three above conditions, the distance function of the visual space must have the two properties: (*a*) If P_1, P_2 and P_3, P_4 are two pairs of light points and an observer reports that the distance between P_1 and P_2 appears greater than the distance between P_3 and P_4, then $D(P_1, P_2)>D(P_3, P_4)$; and (*b*) if an observer reports that the points P_1, P_2 and P_3 appear to lie on a straight line, then $D(P_1, P_2)+D(P_2, P_3)=D(P_1, P_3)$, and conversely.

The distance D satisfying both sets of conditions is the psychometric function Luneburg seeks. The problem is not only to determine this function but also, if possible, to express it in terms of the physical coordinates of the stimuli.

It may be shown that a distance function is determined only up to a linear transformation. That is to say, if $D(P_1, P_2)$ satisfies all the above conditions, then the function $\bar{D}(P_1, P_2)$ does also where, $\bar{D}(P_1, P_2)=C\cdot D(P_1, P_2)$, C being an arbitrary constant. This arbitrariness, the only one possible, means that there is no absolute sensation of size.

For convenience in representing the relationships between visual and physical space Luneburg introduces a bipolar coordinate system. In a cartesian coordinate system let the centers of rotation of the two eyes, R and L, be at the points $x=z=0$, $y=\pm 1$. If the eyes are fixated at a luminous point P^*, the optical axes of the eyes lie along RP^* and LP^*. Let the angle of elevation of the plane containing R, L, and P^* be represented by θ, angle LRP^* by α, and angle RLP^* by β. (See Fig. 1 of the reference of footnote 2.) The bipolar latitude $\phi=\frac{1}{2}(\beta-\alpha)$; and the bipolar parallax, $\gamma=\pi-\alpha-\beta$. (See Figs. 2 and 6 of the reference of footnote 2.)

The problem remains to find the relation between the physical space and the visual space, in which we may establish a cartesian coordinate system ξ, η, ζ, where the ξ, ζ-plane represents the subjective median plane, the ξ, η-plane the subjective horizontal plane, the η, ζ-plane the subjective frontal plane, and the point $\xi=\eta=\zeta=0$, the apparent center of observation.

For infinitely distant points in the physical space, $\gamma=0$. Since these points appear visually to form a dome of finite radius, they are represented in the ξ, η, ζ-space by a sphere concentric to the origin. Arbitrarily a sphere of radius 2 is chosen to represent physical infinity.

If a number of equidistant light points arranged on a Vieth-Müller horopter circle are observed, they will approximately appear to lie at equal distances on a circle concentric to the center of observation. A number of Vieth-Müller circles may now be constructed in the same way at various distances from the eyes. If this configuration is rotated about the y-axis, a Vieth-Müller torus is obtained, giving the impression of a sphere concentric to the center of observation. Expressions for points on this sphere are $\xi=f(\gamma)\cos\phi\cos\theta$, $\eta=f(\gamma)\sin\phi$, and $\zeta=f(\gamma)\cos\phi\sin\theta$; these functions relate visual and physical coordinates. The unknown function $f(\gamma)$ is a non-negative function of γ which increases monotonically if γ decreases from positive values to $\gamma=0$. Because of the arbitrary choice, $f(0)=2$.

Measures of $f(\gamma)$ obtained by means of Vieth-Müller circles[2] show that $\log f(\gamma)$ is a linear function of γ and therefore, $f(\gamma)=2\epsilon^{-\sigma\gamma}$, where σ is a parameter having different values for different individuals. Expressed in bipolar coordinates, Riemann's line element can be written

$$ds = \frac{2}{\epsilon^{\sigma\gamma} + K\epsilon^{-\sigma\gamma}}\,(\sigma^2 d\gamma^2 + d\phi^2 + \cos^2\phi\, d\theta^2)^{1/2}.$$

"The constants σ and K are personal constants of the observer, . . . σ . . . determines the sensitivity of depth perception as compared with the perception of lateral size," while K determines the relation between visual estimation and physical size.[1]

The Hillebrand-Blumenfeld alley problem provides a method for evaluating σ and K. The Hillebrand problem may be described as follows: The subject arranges two rows of light points so that they seem to form parallel alleys symmetrical to the median plane. The Blumenfeld procedure requires the subject to arrange two rows of light points (in darkness) in such a way that all apparent distances between frontally coplanar points are equal. The settings made by the subject in these two situations form diverging curves that are not physically straight. This finding is not accounted for by a theory of corresponding points.

A method of determining σ and K is given by the fact that distance and parallel alleys become asymptotic at infinity. The equation describing the asymptote is $y=\tan\phi_0\,(x+b)$; where $b_D=2\sigma(1-K)/(1+K)$ (Blumenfeld); and $b_P=2\sigma(1+K)/(1-K)$ (Hillebrand). Therefore, $4\sigma^2=b_D\cdot b_P$ and $b_D/b_P=(1-K)^2/(1+K)^2$. This method of determining K is of course independent of any assumption regarding K's being negative, i.e., K can be found in this manner without any assertion of a hyperbolic space.

The qualitative differences between the parallel and distance alleys demonstrate the hyperbolic nature of the visual space, but the method described is not convenient for measuring the parameters of the metric because the experimental technique is usually too elaborate.

Another estimation of σ may be made by observations of apparently right angles.[1] From the line element relation it follows that two line elements $d\gamma$, $d\phi$, $d\theta$ and $\delta\gamma$, $\delta\phi$, $\delta\theta$ attached to the same vertex will be discriminated as perpendicular if and only if $\sigma^2 d\gamma\delta\gamma+d\phi\delta\phi+\cos^2\phi\, d\theta\delta\theta=0$. If two such line elements are determined by experiment,

$$\sigma^2 = -\left(\frac{d\phi}{d\gamma}\,\frac{\delta\phi}{\delta\gamma} + \cos^2\phi\,\frac{d\theta}{d\gamma}\,\frac{\delta\theta}{\delta\gamma}\right).$$

A better method for determining the parameters makes use of single observations of a pair of Vieth-Müller circles. The subject is instructed to equate the distance between the two comparison points on the near circle until it equals the distance between the standard points on the far circle. The method of evaluating the parameters is discussed in Luneburg's posthumous paper.[2]

Luneburg's theory provides a mathematical formulation which may be valuable in systematizing many types of observations.[3] It is to be emphasized that the question of its value will depend upon the adequacy of its predictive power, a power that can be ascertained only when appropriately designed experiments have been carried out.

As it stands, the theory is essentially a formal account based on a mathematical model. It does not "explain." In fact, Luneburg disclaims any desire to account for the "mechanisms" of binocular spatial discriminations. The psychologist, therefore, can go on analyzing the determinants of space perception without necessarily relating his results to the Luneburg theory. Presumably, in the distant future a theory of space will exist that will take account of the (by then) demonstrated "mechanisms" of space perception. Such an account will provide us with a theory in terms of hypothetical constructs.[4] Luneburg's theory is not of this sort; his visual space is a mathematical intervening variable.

---

[3] Some calculations by Stein, based on an analysis that probably stems from Luneburg, may be important for the problem of visual "equivalence," e.g., in "distorted" rooms. Cf. A. Stein, A certain class of binocularly equivalent configurations. *J. opt. Soc. Amer.*, 1947, **37**, 944–962.

The comment that Luneburg's theory is incomplete in failing to take account of distance cues other than stereoscopic factors is irrelevant, albeit true. Luneburg's theory is not a general theory of vision but a restricted one. To test the therory in an experimental situation the effect of other cues must be minimized—optimal viewing conditions cannot be provided.

Luneburg's concept of visual space holds considerable interest for systematic psychologists. There seems to be little doubt that Luneburg conceives visual space to be a phenomenological construct and the question arises: From the point of view of behavior study, what is the visual space to which Luneburg alludes? Very obviously what Luneburg's theory does is to relate some "setting" made by a subject (in the dimensions of Euclidean space) to a controlling magnitude of standard stimulus (also measured in Euclidean space). This type of experimental situation is describable by a formal relation of discrimination,[5] $a_e = f(b, R_1, x_1, y_1, z_1)$ where a_e is the "equality" distance setting; b the standard distance; R_1 the reference response frequency; x_1 the constant effect of instructions; and y_1 and z_1 "conditions" of the organism that are essential ("mechanisms"?) to the discrimination. Thus, Luneburg's theory is a set of mathematical functions providing appropriate transformations to relate subject's "settings" to standard stimulus variables. It says nothing of "conditions" of the subject.

The Ohio State University. P. RATOOSH.
Columbia University. C. H. GRAHAM.

McCLELLAND, DAVID C. *Personality.* New York: William Sloane Associates, 1951. Pp. xvi+654. $5.50.

It is puzzling to know what to say about this book. It has some woeful faults, and yet it makes some fine contributions. There are some flashes of keen thinking and some brilliant writing at a few points, as in the characterization of Freud in Chapter 1 and in some criticisms of Murray's concepts. The book gives a good critical discussion of Gordon Allport's ideas (though the criticisms sometimes side-step the questions raised by Allport); and it even more successfully builds on, and discusses critically, the over-all approach of Murray. A wealth of valuable research material is reviewed, both on animals and humans. A fair part of the latter is taken from studies by McClelland and his associates—studies which, to date, have not been adequately used elsewhere. Chapter 12, on experimental studies of motivation, gives the best critical discussion of motivation that I know of, even though I cannot agree with McClelland's final conclusion that only goal-expectations or anticipatory goal-reactions serve as motives.

---

[4] MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.

[5] GRAHAM, C. H. Behavior and the psychophysical methods: An analysis of some recent experiments. *Psychol. Rev.*, 1952, **59**, 62–70.

McClelland has tried to give a theoretically-oriented rather than a practically-oriented discussion of personality, partly because (as he claims, p. xi) "theory must always precede application." To this end, he puts most emphasis on experimental studies (in a broad sense of the word), and a good deal of emphasis on TAT material (but not on most of the questionnaire-type personality tests because " . . . they are difficult to analyze in terms of any meaningful theoretical units," p. 186). Rather frequent quotations are made from a few clinical workers, particularly from Flügel as a spokesman for psychoanalysis—he has no aversion to using clinical theory as a source of hypotheses, at least.

The organization of the book is as follows: Chapter 1 outlines a thesis as to how mankind developed a faith in the possibility of a science of personality. It is an intriguing discussion, but "thin." Three next chapters deal with the methodology of research and theory-building. Three next chapters deal with "traits," which he describes at different points in rather different terms. Thus, sometimes traits are spoken of as "the characteristic way" a person has learned "to adapt to certain recurrent problems" (p. 156), and as " . . . consistencies in behavior in similar situations under similar motivation" (p. 230). Thus conceived, traits might include some rather fundamental characteristics, and McClelland might have discussed, as traits, the individual's ways of dealing with anxiety (per Freudian notions), and he ought to have described, say, Horney's hypotheses regarding the main patterns of social adjustment. But these Freudian concepts are described (at some length) only in a much later chapter, and Horney's type of hypothesis regarding traits does not appear in the book. Apparently there is a lot of truth in his statement that he " . . . decided for reasons of convenience to use the trait variable *to describe the surface or stylistic manifestations of personality only*" (p. 233; italics his).

Three next chapters discuss "schemata" (conceptions) and their development in the individual. Unfortunately, these chapters are in reverse order from their chronological reference to the individual's life. For class use, I would begin with Chapter 10 on learning problems of early childhood, then take Chapter 9 on roles and role models, and then Chapter 8 on the individual's learning of cultural ideologies.

The book ends with three chapters on motivation, one on the individual's conception of self, and a final chapter devoted mainly to a further linking of the concepts of the book to an illustrative case used at length throughout the book.

As I said earlier, there are some very unfortunate features of the book. Let us start with a minor but troublesome one: There are no chapter summaries, no analytical table of contents, and (despite the title of Part V) no real summary chapter. What is more, the section headings within chapters are often poorly planned—supposedly coordinate points are not logically coordinate, etc. It is extremely difficult, in consequence, to get a clear idea of the main concepts of the book.

From the standpoint of use of this book as a textbook, one chief trouble is that the weakest and dullest part (Chapters 2–7 on methodology and traits) comes at the start of the book. Research is not actually done as he says, "obtaining the facts" first, "interpreting the facts" next, etc. What is more, research on personality has to deal with some grave problems which these chapters leave untouched. Later on, McClelland often stresses the idea that a person typically does not know many of his own characteristics and cannot report many of his personality processes. But there is nothing in the chapters on methodology to suggest that the important facts about personality do not just lie around on the surface of people's lives—that individuals often conceal, misrepresent, and have strong repressions which prevent them from revealing facts even when (in a sense) they want to cooperate with research workers. Indeed, it is significant that the rather full index of the book does not list at all the concepts of resistance, repression, or the Rogerian concept of self-defensiveness. It is true that the concept of repression is discussed under the term "withdrawal" in pp. 508–509, but the methodological sections have no benefit from such notions. The discussion of traits, on most scores, is very unsatisfactory. Any teacher using the book would be wise, I think, to skip these six chapters. I doubt whether a class could be resuscitated after struggling through this early part of the book.

Another fault is that the material is sometimes rather poorly digested, logically speaking. For instance, after trying to build a definition with some care, McClelland says: " . . . we have concluded with a definition of personality as the most adequate conceptualization of a person's behavior in all its detail that the scientist can give at a moment in time. . . . a person's personality . . . may change as he changes or as the scientist's insights improve" (pp. 69–70). Naturally, he doesn't stick to this definition. Such statements bespeak pseudo-sophistication —pretentiousness—not real thinking.

McClelland's concept about the place of motivation in personality is an unduly hampering one. Thus, he says: " . . . we need two concepts—one which will account for the consistencies and recurrences, and one which will account for the inconsistencies and sudden, irrational changes in behavior. For this purpose . . . we may, following Murray, choose the terms *trait* and motive" (p. 215). Similarly, at the end of the book, concluding his discussion of his long illustrative case, he says: "Perhaps 95 per cent or more of his responses . . . can be best accounted for in terms of his internalization of culture patterns (schemata) and his consistent modes of adjustment to recurrent problems (traits). Motives get to be decisive only in a small number of cases, which tend to get blown up beyond their real importance because they are unusual and dramatic" (p. 618). This view it regrettable. The chapters on motivation contain some of the best material in the book, but McClelland hardly has a framework in which he can use this material!

The chief foundation of personality concepts, McClelland says, should be our concepts of learning. But too often the learning concepts are half-understood ones (e.g., that Jost's laws signify that earlier learning leaves more enduring habits than later learning, pp. 341–342; and that "Extinction, as it has been observed in the laboratory . . . is so temporary that it could hardly be considered a serious reason for the disappearance of a trait like expansiveness" p. 219).

The long illustrative case of Karl, used in detail throughout the book, was not a case studied by McClelland, but was reported by one of his students on the basis of an autobiography, Strong Vocational Interest answers, Rorschach and TAT responses, perhaps one interview by a psychiatrist, answers to a sentence completion test, and a few other such things. The clinical workers are often berated by McClelland throughout the book. One could hardly blame them if they retaliated by criticizing this case as evidence of the shallowness of "scientific" studies of personality.

University of Oregon. ROBERT LEEPER.

EELLS, K., DAVIS, A., HAVIGHURST, R. J., HERRICK, V. E., & TYLER, R. *Intelligence and cultural differences.* Chicago: Univer. of Chicago Press, 1951. Pp. 388. $5.00 (lithoprint).

First a word about the multiple authorship of this volume, 90 per cent of which was written by Eells. Herrick's name appears because of a six-page inadequate review of the literature. Both Davis and Tyler get in by rehashing previously published articles, and Havighurst's contribution is six pages. Thus future bibliographers are burdened unnecessarily with four extra names.

Eells reports herein his doctoral research on the relationship of cultural status to total scores and to performances on individual test items. Over 2200 nine- and ten-year-old pupils were tested on the Henmon-Nelson, Kuhlmann-Anderson, and Otis Alpha (verbal and nonverbal); and over 2400 thirteen- and fourteen-year-old pupils were tested on the Terman-McNemar, Otis Beta, Thurstone Spatial and Reasoning (from Primary Metal Abilities battery), and California Mental Maturity. These samples include practically all pupils of the given ages in a city of about 100,000 population. Social status was determined on the basis of an index giving equal weight to parental occupation, parental education, house type, and dwelling area.

As expected, the relationships of total scores to the social status index were positive (correlations of .20 to .43), with the verbal tests showing the higher relationships.

The unique part of this research is the extensive item analysis, which is concerned mainly with the performance of a high-status group (top 10 per cent of the status index distribution) compared with that of a low-status group (bottom 15 per cent of distribution). The N's for these comparisons are about 225 and 325 for the lower age group and about 235 and 360 for the higher age group. Of the 658 items involved in this analysis, 460 showed differences in percentages (passing) significant beyond the $p = .01$ level while 153 did not reach the .05 level of significance.

The chief finding is that the verbal type items, in general, yield the larger differences. Status differences in passing items are not dependent either on the position of the correct response or on type of question (analogies, classification, etc.). Status differences are related to item difficulties for verbal items, but for nonverbal items no generalization seems possible.

Two long and tedious chapters are devoted to a comparison of high- and low-status groups as regards correct and incorrect responses to certain items. Here one finds after-the-fact rationalizations in terms of differences in cultural backgrounds, but no explanation is found for a "rather substantial number of items" showing large status differences. There seems to be some dismay that certain supposedly culturally biased items did not yield status differences.

The study, though rather well executed and reported, is not free of questionable points. Pertinent differences in test SD's are not considered (pp. 146–147; 155). The correlational term is ignored in computing certain standard errors of differences (pp. 114–117; 136–139). Subgroups are compared statistically with total groups without necessary allowance for lack of independence (p. 191). Inflated N's are encountered (pp. 201–204).

Eells, perhaps in tune with his mentors, concludes that "variations in opportunity for familiarity with specific cultural words, objects, or processes required for answering the test items seem . . . to be the most adequate general explanation for most of the findings." He favors the construction of a common-culture test as a means of securing scores not subject to cultural bias. There are those who will say that Eells has not yet performed the crucial experiment since possible (and likely) genetic factors were uncontrolled.

QUINN MCNEMAR.

Stanford University.

GURVITZ, MILTON S. *The dynamics of psychological testing*. New York: Grune & Stratton, 1951. Pp. xii+368. $6.75.

This is a disappointing book. Its purpose is excellent: to present and analyze seventeen unselected cases which represent the routine work during one month of a clinical psychologist in a mental hospital. For each case we are given the diagnostic problem presented, complete protocols of all tests (including reproductions of figure drawings), analysis of the protocols, the psychologist's report (written before the case history was seen), the psychiatric therapist's comment on the validity of the statements in the psychologist's report, case history, and a final summary critique. There is an opening statement on the clinical use of tests which this reviewer found admirable and stimulating. The author then presents the discussion of the tests used. His comments on the Wechsler-Bellevue are adequate, but one often wishes that the field offered him more experimental data to back up some of the statements made. The sections on the Rorschach are disappointing. They are written much too briefly, on a level which will not be intelligible to the beginner, and do not offer enough discussion and elaboration for the experienced worker. The chapter on diagnostic formulations is sketchy and inadequate.

The case records are complete and interesting and will make excellent teaching material. The comments of the psychiatrist on the psychologist's report are interesting, but again sketchy. The flat acceptance of a single psychiatrist's statement of agreement or disagreement as an adequate criterion for the validity of the psychologist's interpretations is of dubious objective value. When these agreements and disagreements are counted and the chi-square formula is then applied, the result is an impression of statistical pretentiousness completely unwarranted by the inadequacies of the data. Scattered throughout are many pithy statements and illuminating comments, but the book as a whole suffers from loose organization. The publishers might well have supplemented the excellent format they have given the book by a tighter editorial rein on the author.

Northwestern University. WILLIAM A. HUNT.

JANIS, IRVING L. *Air war and psychological stress: Psychological studies of bombing and civilian defense*. New York: McGraw-Hill, 1951. Pp. x+280. $5.00.

This scholarly book is an eery sign of the times. In studies originally undertaken for the RAND Corporation and published under its *imprimatur*, Janis winnows the meager and inconsistent testimony from Hiroshima and Nagasaki and the richer though often impressionistic literature on civilian response to heavy bombardment in England, Germany, and Japan for their yield of empirical generalizations about psychological reactions to attack by the "conventional" and "unconventional" weapons of modern air war. In his competent execution of this essentially historical task, he helps to secure for psychology the extended knowledge of human potentiality for adjustment that is a grim by-product of total war. From his critical survey of the evidence he then draws tentative conclusions about the underlying psychodynamics that should be of considerable interest to students of personality. The final sections of the book are concerned with psychological problems of civilian defense against atomic attack, partly by way of tentative policy recommendations, partly in terms of suggestions for needed research. The book is cast in a qualified, academic tone that contrasts with the appallingly unacademic facts of the past and assumptions about the future with which it deals.

Examination of published eye-witness accounts, together with a re-analysis of the original interviews from Hiroshimo and Nagasaki conducted by the Morale Division of the U. S. Strategic Bombing Survey, leads Janis to reject the allegation that the atomic bomb produced any widespread panic, disorganized activity, or antisocial behavior. With the difference that all atomic survivors had experienced a "near miss"—rather than the minority as in the case of ordinary bombing—both the immediate and the more enduring psychological effects of atomic bombing seem to have been surprisingly similar to those of heavy conventional bombardment. The more adequate evidence available for conventional bombing thus becomes pertinent for assessing the factors that may influence psychological response to atomic attack.

The facts in broad outline about civilian adjustment to air war are generally known and accepted. Janis summarizes them thus:

> Mental breakdown, panic, and mass demoralization—the triple psychological threats that dominated so much of the thinking in official quarters—rarely materialized during World War II. . . . the dire predictions made by many self-styled "experts" on mass behavior failed to take account of the psychological stamina of the average citizen. Bombing had little effect on the incidence of mental disorder; outbreaks of mass panic were rare; behavioral morale was maintained at a relatively high level. Although emotional shock did occur on a sizeable scale, most cases recovered fairly rapidly.

The details are more controversial, and Janis weighs judiciously the available evidence on such issues as the relative effectiveness on morale of distributed vs. saturation bombing, the apparent diminishing returns of the heaviest bombardment, and the predominance of intragroup vs. extragroup aggressive reactions.

Janis finds considerable support for MacCurdy's hypothesis that it is the experience of being a near-miss rather than danger as such that produces emotional disturbance; without near-misses, fear reactions tend to diminish during a series of air attacks. Such experiences, he offers in explanation, undermine the illusion of personal invulnerability which gives effectiveness to many of the defense mechanisms for coping with anxiety.

Unlike the sections of the book that deal with the psychological effects of bombardment in World War II, which are well organized and comprehensive in their treatment, the concluding chapters on problems of civil defense against atomic attack are in the very nature of the case more sketchy. Among the problems treated are disaster control, recruiting and training civil defense volunteers, mass education for individual survival, and coping with apprehensiveness in potential target populations. The prime value of these chapters, which remain perhaps too close to the research memorandum stage for fullest usefulness to the general reader, seems to the reviewer to lie in their alertness to the complexity of the psychological problems involved in decisions of practical policy. When does "inoculation" to the horrors of atomic disaster inoculate, and when does it merely demoralize? How can civil defense teams be effectively motivated to train for disaster control work in neighborhoods or cities other than their own (for the teams in the target neighborhood cannot be expected to survive)? The problems, practical and theoretical, are immense, and they will be grappled with more intelligently on the basis of such considerations as Janis introduces.

Vassar College. M. BREWSTER SMITH.

ROGERS, CARL R. *Client-centered therapy; its current practice, implications, and theory.* Boston: Houghton Mifflin, 1951. Pp. xii+560. $4.00.

This volume covers the ten-year period since Rogers' *Counseling and Psychotherapy* was published. Besides copious material on individual cases and on studies of them, it has chapters on play therapy, group therapy, leadership and administration, and teaching, by members of the school who have had special experience in those applications of the therapy. There is also a chapter on a theory of personality and behavior, by Dr. Rogers.

Very early in the book there is a presentation of two striking and radical changes in point of view from the earlier expositions. The effectiveness of client-centered therapy is now said to be, first of all, a matter of the therapist's character: he must have a system of attitudes that is described at some length. The system of attitudes amounts to a profound belief in the sufficient (or unlimited?) capacity of a client to understand and reorganize himself on his own responsibility, not the therapist's. The problem of therapy then becomes: what is the best way of "implementing" these attitudes? Second, the counselor's role is now to convey understanding of attitudes and feelings. More variety in attempts to do this is anticipated, to the point where the current formulation of the best way may be discarded. The current formulation, as I understand it, is that the therapist concentrates his whole effort on achieving a deep understanding of the private world of the client, and uses only responses that are attempts to convey understanding, excluding all others as far as possible, and conveying nothing deeper of his understanding than what the client perceives at that point. Although the door is opened to others, this is the only "implementation" treated to any great extent in the body of the book. It is curious that in a school which has prided itself on its vigor in research, and been admired for that, these striking changes should not have come as a result of its research. Yet they have not; they are laid to clinical intuition and experience. It is suggested that the experience was that cases grew longer along with the changes, and this involved an increasing degree of personality reorganization.

As I read along in the examples, illustrations, and citations of research, I was left feeling that there was a serious lack. The source of this came to light on pages 188 and 189, where I read of "our inability to profit in a research way from our failures," and "We have not been able to learn, in any general or significant way, from the clients whom we fail to help." I asked, "Then what is happening here?" We are told that a method works. We are shown devices accepted as legitimate samples of the method because they work. We would assume, then, that without them you get failures, and that with them you don't get failures. But it appears that the study of failures does not show this. Then what does "work" mean?

For that there are two tests: (1) The client's responses during the sessions. For example, a response shows insight, self-acceptance, further confession, and the like. This test is used, and goodness or badness of counselor activity is evaluated accordingly. (2) The client's success or failure in being helped. This one is not used, it fails to discriminate. That is, counselor activity (or client response?) cannot be shown to be different in the two. It would follow that the client's responses during the sessions are related to counselor activity, but that the client's success or failure in being helped is not related to counselor activity, or is not related to the client's responses during the sessions. I am not sure whether it is one or the other, or both. Should not the two tests be expected to have something in common, if they are to have importance in the long run? Although successes are said to be due to the use of non-judgmental speech and expression and of conveyances of understanding, failures are not due to their absence, or to their opposite. Until we know under what conditions a method fails, do we really know anything about why it succeeds when it does succeed? How else do we ever discover the illusory elements in our "successful" methods of doing anything? After all these pages, we seem to come out where we have been before. Interest in the great mass of material begins to fade.

It is easy to be annoyed when people demand proof from this school of a kind that they never demand from their own, but this school has in a way tried to meet that challenge in its research. There is much to admire and agree with in the discussions of general and vitally important problems in the book, but essentially the book is an argument for a very particular method. It is the evidence on this that is disappointing.

Western Reserve University. DWIGHT W. MILES.

WELFORD, A. T. *Skill and age. An experimental approach.* London: Oxford Univer. Press, 1951. Pp. ix+161. 8s. 6d.

The significance of the book by Welford and his associates is more in the development of a point of view and in methodology than in the actual findings presented. Thus the observation that with increasing age, there is a gradual reduction in performance of tasks involving "time-stress" will come as no surprise to those familiar with the experimental literature on aging—or to those familiar with old people. However, the emphasis on determining the relationship of age to the manner in which achievement is attained is of great significance. Furthermore, the presentation differs from most studies on aging in that a conceptual theory is developed, predictions are made on the basis of the theory, and results of experiments designed to test the theory are reported.

Although the importance of physiological factors is recognized, major emphasis is placed on the psychological and environmental aspects of the problem of age changes in performance. Thus, the differences between young and old in their methods of dealing with situations and problems is analyzed in some detail. It is postulated that the central factors of organization both of sensory data and response mechanisms are of primary importance in the diminished performance of the elderly.

The problem of sampling in the study of age changes is well recognized by the author and is met in a forthright manner, that is, by testing subjects he could get. Most of the subjects under 30 were students or naval ratings taking part in Service experiments in Cambridge. Subjects from 30 to 80 years of age were: (*a*) friends and acquaintances, (*b*) friends, tradesmen, and acquaintances recruited by a relative of one of the members of the research staff, (*c*) army officers obtained through the Cambridgeshire and Isle of Ely Territorial Army Association, (*d*) Polish army personnel and agricultural workers obtained from two Polish Resettlement Corps camps and two County Agricultural Executive Committee's Polish Hostels, (*e*) adult education classes and (*f*) members of old people's clubs. Although it is stated that age groups were equated with respect to socioeconomic status, education, and occupation, the evidence is not presented.

The experiments consisted of tests involving simple motor skills (such as tossing rings at a target), more complicated apparatus tests (tracking tests), tests requiring reversal of practiced skills (mirror drawing), tests of logical thinking on inspection tasks, and the solution of electrical problems. Many of the tests used were most ingenious and all were designed so that many partial scores, indicating the methods used by the subject in reaching a solution or completing a performance, could be obtained. Unfortunately the apparatus used is not described in sufficient detail to permit duplication by others.

Investigators hoping to find detailed observations on the regression of various test scores on age will be disappointed, since most of the results are analyzed in terms of differences between three to five broad age categories. The number of subjects tested varied from experiment to experiment, ranging from 12 to 75 per age group. The age groups compared also varied from "above 30" and "under 30" to 10-year intervals.

Some of the known psychomotor characteristics of old people are confirmed, such as the slower, more deliberate and accurate performance of older people, their greater susceptibility to confusion, their greater variability in test scores. However, the improved methodology of this study makes possible the development of a new concept, viz., that the impairment in performance in older people is not due only to peripheral, sensory, and motor impairments, but also to alterations in central factors concerned with the perceptual organization of sensory data and perhaps the central organization of responses. This is a valuable concept which will require more detailed investigation.

In addition to the theoretical implications of this study, there are a number of findings of immediate practical value. For instance, there is an indication that in training older workers, written instructions may be much more efficient than demonstrations. It is also evident that many well-learned and practical skills may be performed by older persons with as great efficiency and accuracy as the young. It is also suggested that older workers may perform more efficiently in groups of workers of their own age than when included among groups of younger workers.

The approach and methodology developed by Welford and his associates deserves careful consideration of all students of aging.

NATHAN W. SHOCK.

Baltimore City Hospitals.

BERRIEN, F. K. *Comments and cases on human relations.* New York: Harper, 1951. Pp. xi+500. $4.50.

Inspired by contact with Dean Wallace B. Donham and his well-known Harvard course, Berrien has prepared a worth-while book on problems in human relations. Its level and content have been determined partly by the author's experience with a collaborative human relations course at Colgate University. It is a textbook pitched at the level of college students and adult education classes; it is not an abstract, theoretical contribution which the publishers have called a text in order to promote sales.

The book's thesis is that "friction and discord in human relations can be reduced provided (1) we develop a clear understanding of the context and core of our conflicts in realistic rather than abstract terms; (2) we carefully assay the alternative lines of action in particular instances; (3) we attempt to predict the consequences of each alternative." Development of an explicit theory of human relations, says Berrien, is not an essential condition to improving human relations. However, "there are certain discoverable generalizations which describe the conditions underlying cooperative, satisfying group behavior"; these can be discovered through psychological and social science investigation.

Part I, the first half of the book, consists of "Comments" centering about aspects or conditions of behavior which affect social relationships. Chapters on "words and things" and on observing and thinking stress the importance of subjective interpretations, of inferences and generalizations, as affecting interpersonal relations. A section on motivation emphasizes the self-actualization drive. Frustration of this or other patterned motives may produce either aggressive or apathetic reactions. The importance of social skills and group codes and the problems caused by conflict between individual and group values are noted.

Part II consists of 28 case studies, running from those which emphasize the individual to those involving, in greater degree, relations between groups. These are actual life stories from a variety of settings—home, school, club, college, job, marriage. They typically end on an unfinished note, which facilitates discussion and suggests the need for more information. An instructor's appendix is designed to aid the teacher in taking a clinical point of view toward human relations problems.

Human relations courses seem to be taught by an extremely varied group of instructors; their emphases might not coincide with those of the author, who is known as an applied psychologist. I for one would have preferred more emphasis on emotional factors and on social roles as well as codes. Or again, I would favor stressing personality trends in leaders fully as much as in followers. But the bringing out of such differing interpretations is part of the task of class discussion, and an important aspect of the teaching process.

Curiously enough, Berrien does not indicate very clearly his frame of reference for interpreting problems in human relations. As already indicated, his viewpoint is not narrowly psychological. At times, however, the reader is puzzled. For example, the author makes a distinction, on the basis of complexity, between the kind of cases handled by social agencies and clinics and the kind he presents in the book. "The disturbances of human relations in everyday life," he says, "are relatively simple and consequently are amenable to interpretation without the specialized, detailed information on mental level, endocrine balance, etc., that is essential for a clear understanding of the more puzzling and baffling problems confronting the clinic and the hospital." The relation between the two types of case, Berrien continues, is analogous to the relation between a doctor and a first-aider. "The situations dealt with in this course are of the first-aid variety. If properly handled, they will not need professional care. If bungled through ineptitude, they may fester into more serious problems" (pp. 466–467).

To me, this implies that seriously disturbed human relations problems have an organic basis, or at least derive from personality disorders, whereas everyday garden-variety difficulties are due to cultural or social influences of some sort. Yet a great deal of evidence could be cited to show that much of our divorce, delinquency, labor strife, and war itself stems from technological and economic changes or from various outmoded culture patterns. Furthermore, it is a real question whether everyday disturbances in human relations can be dealt with adequately by students briefly trained in social first aid.

For upperclassmen with a good background in psychology and social science this book may seem too simple. Most beginning students, however, will be stimulated to further exploration in a fascinating but relatively uncharted area.

S. STANSFELD SARGENT.

Barnard College, Columbia University.

ELLIS, ROBERT S. *Educational psychology: A problem approach.* New York: Van Nostrand, 1951. Pp. xi+546. $5.00.

This is a well-written, well-organized book with clearly defined objectives which will meet a definite need in many colleges. Ellis differs from the authors of most recent educational psychology texts in believing that the educational psychologist should be a technical expert but not a policy maker. The psychologist's sole function in education, he believes, is "to point out the methods to be used in order to obtain the desired result with the greatest degree of efficiency." The objectives themselves are set up on the basis of philosophical and social considerations and should not be the concern of the psychologist.

Although the reviewer considers this position completely untenable, the resulting concentration on how to do what the schools are already trying to do has helped the author to produce a book which will be very valuable to teachers and to teachers-in-training. The problems referred to in the subtitle are dealt with under the following topics: the admission and classification of pupils, the curriculum, extracurricular activities, methods of teaching, the development of attitudes, tests and examinations, study methods, mental hygiene, exceptional children, counseling and guidance, discipline, and the problems of the teaching profession.

Since four and five years of professional training are becoming common there is a growing feeling that the teacher should be a professional in the most complete sense, that he should have such a clear understanding of human nature and of educational objectives that he can vary his practice with the situation and deal with situations which have never arisen before. Quite often, in our efforts to reach this ideal, our students acquire a fine inventory of psychological principles which they cannot quite connect with their specific teaching practices. Because he starts with quite definite problems Ellis, however, is able to work down to very specific applications. The following sentence, taken at random from the chapter on developing concepts, is an example of the detailed level of practice at which he consistently arrives: "In the case of both sound recordings and moving pictures there is one very important caution: it will be generally advisable that they be repeated several times in order that the desired observations may be more nearly complete and accurate."

The problems are discussed with fairness, good humor, and a conspicuous lack of dogmatism. Traditional as well as progressive procedures are given their due. The author is not one to throw out the baby with the bath; on one or two occasions he seems willing to leave a little dirt in the tub just to be safe.

His own point of view is consistent and helps the student to relate the various problems to one another. It is holistic, with the stress on meaningful learning as the chief educational process.

The topics are well documented, the relation between psychological theory and teaching methods is well developed, and the figures are not distorted to make a point. Although it may seem too good to be true, there is only one graph in which the zero point on the vertical axis is not shown.

Although this book could be used by a class which had no previous training in psychology, the author has intended that it follow a course in general psychology and a course in child psychology. It should be most effective and very helpful when used in that sequence and just before or just after student teaching.

DONALD SNYGG.

State University of New York, Teachers College, Oswego.

ENGLISH, HORACE B. *Child psychology*. New York: Henry Holt, 1951. Pp. xiv+561. $4.50.

In 1941 appeared a paper-bound volume, *Studying the Individual School Child*, by Horace B. English and Victor Raimy. This was a unique contribution to the field of texts in child psychology. It was designed as a sort of guide to the student actually studying the child—a real child in his real environment. English and Raimy gave solid suggestions as to what to look for, how to go about getting essential information and, finally, how to evaluate what you have when you are through.

To this reviewer it is apparent that English has extended this general philosophy, developing now a book in child psychology. In fact, much of the meat of the earlier volume with Raimy has been incorporated bodily. *Child Psychology*, in this case, is a text developed around the idea that the student needs to work with a child, or children, and needs guidance in doing so. As English puts it in his preface, "If the choice lies between having the students read more, even of first-rate accounts of child behavior, or having them observe children, the latter is to be preferred" (p. ix).

From these introductory remarks it should be obvious that this text is not a source book, a bibliographic reference, or an account of what has been done experimentally or otherwise in the field of child psychology. It is a clear, readable presentation of material for the student on how to study children and why it is important to do this carefully. But suggestions are included regarding angles (all too often neglected and of real concern to workers in child psychology today) on the need for selling parents on the idea that *their* child should be studied, and on the special problems of teachers, who may want to be cooperative, but who simply may not have the time (or the principal's support) to give freely.

If you are teaching child psychology, as English would teach it, to people who want to develop a feel for the *child*, this will be an excellent text. If your special need is for an authoritative compendium of research on children, you will find this book unsatisfactory. The author is not unsound (he reveals a real knowledge of the basic literature in the field) but he is dedicated to the position that the student should study the child in the most effective way. If your course in child psychology has this orientation, you will find this a valuable text.

T. W. RICHARDS.

Northwestern University.

HORROCKS, JOHN E. *The psychology of adolescence.* Boston: Houghton Mifflin, 1951. Pp. xxvi+614. $4.50.

This new textbook on adolescent psychology clearly makes an outstanding contribution to the literature in this field. The book is unusually well documented and draws heavily upon research studies. The 1,500 references which are cited not only include some of the best of the older research, but in addition cover in a very thorough manner the newer research of the past ten years. The book, however, is not a mere compilation of research studies. Horrocks has written a very thoughtful book and has succeeded in developing generalizations and an integrated viewpoint which is all too uncommon in texts in this field.

A particularly noteworthy feature of the book is its emphasis upon the social and cultural factors which influence adolescent behavior. Horrocks first devotes six chapters to a treatment of the social psychology of adolescence before giving attention to the more traditional topics of intellectual and physical growth. Such an approach not only points up the fact that adolescence is primarily a social phenomenon, but also tends to make the book more interesting and readable for students. The book has good balance and no significant areas of knowledge are omitted that have relevance to an understanding of adolescent behavior. Unusually complete analysis is made of adolescent activities and interests in later chapters, and the book ends with two very valuable summarizing chapters. The first of these contains an integrated theory of human adjustment and the second is an extended and penetrating case study of an adolescent.

In a book as comprehensive as this one, certain errors inevitably creep in. For example, it is stated that "Kretschmer also proposes a deviate type which he calls athletic" (p. 292). Kretschmer's special or deviating types were instead designated "dysplastic." The atheltic type represented one of his three major categories of body build. On page 289, McCloy is quoted, but the quotation is not contained in the reference which is cited. In general, however, the book seems to have been carefully prepared and to be relatively free of such faults.

In the main, Horrock's book is highly objective and modern in its approach and should do much to put the study of adolescence on a scientific basis. It furthermore is molar and functional in its outlook. Psychologists, counselors, teachers, and others who work with adolescents will find that the book contains much material that can be easily translated into practice. Students who have read the book as a text since the date of its publication have uniformly expressed favorable comments regarding its usefulness, completeness, and readability.

This text, in the opinion of the reviewer, is on the whole a superior production which will be widely used, and which will exert a very wholesome influence on the field of adolescent psychology.

University of Illinois. GLENN M. BLAIR.

SYMONDS, P. M. *The ego and the self.* New York: Appleton-Century-Crofts, 1951. Pp. ix+229. $2.50.

This charming little book is a sequel to *The Dynamics of Human Adjustment*, amplifying the approach to the concepts of the ego and the self that constituted a chaper in the previous volume. In a brief introduction, Symonds reviews some of the problems that arise in dealing with the ego and the self, from Kant to Gardner Murphy, Snygg and Combs, Frenkel-Brunswik, Werner Wolff, and Carl Rogers. That the relationship of the ego and the self is complicated, no one will deny. Symonds proposes to dispose of the confusion of the past by regarding the ego as objective, and the self as subjective. However, as he proceeds, with the ego, the self, the superego, the self-ideal, and the "individual himself," all in the rôle sometimes of agents and sometimes of objects, the confusion between what is subjective and what is objective might lead the critical student into semantical and syntactical problems that would suggest to him that Symonds' solution is not so simple as it appeared to be at first.

The accounts of the structure, functions, and development of the ego and the self; the influence of self-feeling and self-evaluation on behavior; ego strength and weakness; the rôle of affects; the relations between the ego, superego, and self-ideal; ego breakdown and depersonalization; the rôle of the ego in pathological states; the place of the ego and self in psychotherapy; and Symonds' own suggestions about proper education as a means of arriving at satisfactory ego and self development—which constitute the contents of the volume—make interesting reading. Frequent references to concrete examples of behavior, particularly in children, lend an air of hard-headed practicality to even the most involved of the theoretical discussions and should go a long way toward convincing the reader that an individual's concept of himself must be reckoned with in any molar account of his behavior.

The point of view may well be called neo-Freudian, in that Symonds takes over the Freudian ego and superego, but introduces the self as an essential systematic concept. The id is not mentioned, though there is frequent reference to "basic needs and drives." Although Symonds warns against the dangers of reification and states specifically that the ego and the self are *processes*, he personalizes these "processes" so convincingly in his writing that only a relatively critical and discriminating reader could avoid finishing the book under any impression other than that Symonds has made out an excellent case for the belief that human behavior is controlled by the interpersonal relationships of a community of manikins within "the individual himself."

While engaging in polemics, Symonds does so in a kindly spirit that could hardly hurt anyone's feelings. Obviously fond of his manikins and their hosts, he imparts to the reader a kind of good will and tolerance for his fellow man that should be an asset in psychological counseling and psychotherapy.

The annotated bibliography should prove useful to many readers. As to the contention that the book "presents systematically and in organized fashion the results of experimental findings and indications from clinical experience concerning the ego and the self," it appears that perhaps 36 of the 211 titles refer to studies that might reasonably be considered to be experimental in nature, of which 16 concern the level of aspiration and level of achievement. Approximately 20 of the experimental studies are referred to in the book proper. Nothing by Harry Stack Sullivan is included in the bibliography and his theory of the development of the self is not mentioned as such.

ARTHUR JENNESS.

Williams College.

CULBERTSON, JAMES T. *Consciousness and behavior*. Dubuque, Iowa: Wm. C. Brown, 1951. Pp. vii+210. $4.50.

The purpose of this book is to present, from a strictly mechanistic point of view, a neural analysis of consciousness and behavior. To describe the relationship between input (stimuli) and output (responses) of human and animal machines, the author constructs systems of *throughput* in the form of hypothetical nerve nets, which represent patterns of neural interaction. Basic in accomplishing this are postulates of neural transmission and quantized time, standard in the field of mathematical biophysics. These postulates, though overly simplified, allow simple symbolic logic to be used in making statements describing nerve nets and can be easily replaced by the more accurate, but more complex, postulates of experimental physiology.

While his basic approach is quite similar to that of other workers in the field, such as Rashevsky, Pitts, McCulloch, and Landahl, the author diverges rather widely from them in terms of his ultimate objective. Instead of being concerned with establishing general principles of neural interaction, Culbertson becomes more interested in constructing nerve nets which account on this level for specific psychological problems. For instance, he presents nerve nets that could possibly account for the "Gestalt problem," the phi phenomenon, the ability to time events, and the capacity for memory. One area in which he shows even greater divergence from other mathematical biophysicists is that of the mind-body problem. Where such thinkers as Rashevsky refuse to admit a phenomenon of consciousness, Culbertson takes the view that consciousness cannot be ignored and seeks to account for it as a by-product of the interaction of neural impulses as they proceed from receptors to effectors. He thus offers neurological bases for a *psychospace* and for sensory intensity and quality. That these nerve nets do not constitute actual explanations is readily apparent. They are offered, rather, as possible neural connections underlying different types of psychological phenomena. It is then up to the psychologist to specify those characteristics of nerve nets which will be compatible with his empirical observations and thereby help select those that actually obtain in organisms. Once the proper nerve nets have been determined, behavioral characteristics can be further delimited and the psychologist is aided in predicting responses.

The psychologist who does not like to spend hours wading through complicated formulae will be glad to know that this book, unlike most in the field of mathematical biophysics, contains a minimum of logical constructions, and these are quite ably explained in the earlier sections. For such a complex area of study the text is clearly and simply written and should prove fruitful for use in classes in psychological theory or in more advanced classes in physiological psychology.

F. J. MCGUIGAN.

Human Resources Research Office, Department of the Army.

Symposia of the Society for Experimental Biology. IV. Physiological mechanisms in animal behavior. New York: Academic Press, 1951. Pp. vi+482. $6.00.

This volume contains 21 papers read at a Symposium of the Society for Experimental Biology, which was held at Cambridge University in July 1949. The program was subdivided as follows: "The Range and Capabilities of the Sense Organs," 4 papers; "Central and Peripheral Control of Behaviour Patterns," 5 papers; "Instincts, Taxes, etc.," 8 papers; and "Learning," 4 papers.

Of the 22 contributors, representing five different nationalities, two are Americans and only one is a psychologist. Some of the papers are of outstanding quality. Erich von Holst, of the Max-Planck-Institute in Wilhelmshaven reports a series of truly beautiful experiments designed to elucidate the sensory control of positional responses in fishes. B. B. Boycott and J. Z. Young describe an impressive experimental program dealing with neural mechanisms involved in cephalopod learning. K. S. Lashley's stimulating analysis of the current status of various neurological theories of learning serves chiefly, as he concludes, to show "what and where the memory trace is not." It is a considerably more convincing evaluation of current knowledge than is that of J. Konorski, who continues to think of conditioning as resulting from the formation of connections between different brain centers.

The papers vary a great deal in quality, but, taken as a whole, the volume should be extremely useful to American psychologists. Laboratory research workers may feel that some of the investigations carried out by biologists are less thoroughly controlled and less cleanly designed than might be wished. Certainly some of the zoologists display a lamentable lack of sophistication in the conceptualization of psychological problems, and there is an occasional tendency toward philosophical disputation.

Much more significant than any negative criticism, however, is the fact that the study of behavior is being pressed forward upon so broad a front in other countries. It is noteworthy, also, that this work is being done, not by psychologists but by biologically trained investigators. In many respects they are students of behavior in a truer sense than are professional psychologists. Certainly no one could read this collection of papers and escape the conviction that a genuinely comparative approach to psychological problems is fruitful and worth while.

FRANK A. BEACH.

Yale University.

BOOKS AND MATERIALS RECEIVED

Alexander, Franz, & Ross, Helen (Eds.) *Dynamic psychiatry.* Chicago: Univer. of Chicago Press, 1952. Pp. xii+578. $10.00.

Bauer, Raymond. *The new man in Soviet psychology.* Cambridge, Mass.: Harvard Univer. Press, 1952. Pp. xxiii+227. $4.00

Baumgarten, F. (Ed.) *La psychotechnique dans le monde moderne.* Complete report on 9th International Congress of Psychotechnology (Berne 1949). Paris: Presses Universitaires de France, 1952. Pp. xi+ 630. 2.000 fr.

Boisen, Anton T. *The exploration of the inner world.* New York: Harper (1936) reissue, 1952. Pp. xiii+322. $4.00.

Brunswik, Egon. *The conceptual framework of psychology.* Vol. 1, No. 10, International Encyclopedia of Unified Science. Chicago: Univer. of Chicago Press, 1952. Pp iv+102. $2.00.

Bryan, Alice I., & Leigh, Robert D. *The public librarian: A report of the public library inquiry of the Social Science Research Council.* New York: Columbia Univer. Press, 1952. Pp. xxvii+474. $6.00.

Bychowski, Gustav. *Psychotherapy of psychosis.* New York: Grune & Stratton, 1952. Pp. viii+328. $5.75.

Campbell, Clyde M. (Ed.) *Practical applications of democratic administration.* New York: Harper, 1952. Pp. x+325. $3.00.

Carmichael, Leonard, & Mead, Leonard C. *The selection of military manpower. A symposium of the National Academy of Sciences.* Washington, D. C.: National Research Council, 1951. Pp. iii+269. $2.50.

Cattell, Raymond B. *Factor analysis: An introduction and manual for the psychologist and social scientist.* New York: Harper, 1952. Pp. xiii+462. $6.00

Commission on Financing Education. (John D. Millet, Executive Director:) *Higher education and American business.* New York: Baldwin & Mermey, 1952. Pp. 37. (No charge. Copies may be obtained by writing Commission's Ofc., 1800 Broadway, New York 23, N. Y.)

Committee on Education, Training and Research in Race Relations of the Univer. of Chicago. (Louis Wirth, Chairman.) *Inventory of research in racial and cultural relations.* Vol. 4, Bull. No. 2, Winter 1951–52. Chicago: Univer. of Chicago Press, 1952. Pp. 84.

Curran, Charles A. *Counseling in Catholic life and education.* New York: Macmillan, 1952. Pp. xxvi+462. $4.50.

Doob, Leonard W. *Social psychology: An analysis of human behavior.* New York: Holt, 1952. Pp. xix+583. $5.00.

Fitch, Frederic B. *Symbolic logic.* New York: Ronald, 1952. Pp. x+238. $4.50.

von Foerster, Heinz (Ed.) *Cybernetics*. Transactions of the eighth conference (March 1951). New York: Josiah Macy, Jr. Foundation, 1952. Pp. xx+240. $4.00.

Gray, J. S. *Psychology in industry*. New York: McGraw-Hill, 1952. Pp. vii+401. $5.00.

Greenacre, Phyllis. *Trauma, growth and personality*. New York: Norton, 1952. Pp. xii+328. $4.50.

Hall, V. E. (Ed.) *Annual review of physiology*. Vol. 14. Stanford: Annual Reviews Inc., 1952. Pp. ix+589. $6.00.

Kerr, Willard A. *Manual for the tear ballot for industry: A measure of job satisfaction*. Chicago: Psychometric Affiliates, 1948. Pp. 38.

Kerr, Willard A. *Supplement to the manual for the tear ballot for industry*. Chicago: Psychometric Affiliates, 1948. Pp. 8.

Kohler, Iva. *Über Aufbau und Wandlungen der Wahrnehmungswelt: Insbesondere über bedingte Empfindungen*. Vienna, Austria: Rudolf M. Rohrer, 1951. Pp. 118.

Kongressbericht (Bonn 1947). *Charakterologische Schriften des Berufsverbandes Deutscher Psychologen* Hamburg: H. H. Nolke, 1948. 3 Vols. Pp. xvii+125, 145, 143.

Mettler, Fred A. (Ed.) (Columbia Greystone Associates.) *Psychosurgical problems*. Philadelphia: Blakiston, 1952. Pp. xii+357. $7.00.

Morse, William C., Ballantine, Francis A., & Dixon, W. Robert. *Studies in the psychology of reading*. Ann Arbor, Mich.: Univer. of Michigan Press, 1951. Pp. ix+188. $3.00 cloth, $2.50 paper.

Mukerjee, Radhakamal. *The dynamics of morals: A sociopsychological theory of ethics*. London: Macmillan, 1952. Pp. xxvii+530. 25/-.

National Psychological Research Council for the Blind. *Research suggestions on psychological problems associated with blindness*. Washington, D. C.: Federal Security Agency, Ofc. of Vocational Rehabilitation. Pp. 20.

Piaget, Jean. *Play, dreams and imitation in childhood*. New York: Norton, 1951. Pp. ix+296. $5.00.

Piéron, H. *Les problemes fondamentaux de la psychophysique dans la science actuelle*. Paris: Hermann, 1951. Pp. 63.

Piéron, Henri, Pichot, Pierre, Faverge, J. M., & Stoetzel, Jean. *Methodologie psychotechnique*. (Book II in Traite de Psychologie Appliquée.) Paris: Presses Univer. de France, 1952. Pp. vii+339. 1.000 fr.

Pigors, Paul, & Myers, Charles A. (Eds.) *Readings in personnel administration*. New York: McGraw-Hill, 1952. Pp. xii+483. $4.50.

Russell, Bertrand. (Denonn, Lester E., Ed.) *Dictionary of mind, matter and morals*. New York: Philosophical Library, 1952. Pp. xiv+290. $5.00.

SAITO, T. (Chief Ed.) *Folia Psychiatrica et Neurologica Japonica.* Niigata, Japan: Dept. of Neurology and Psychiatry, Niigata Univer. College of Medicine. 1951. Vol. 4, No. 4. $1.50 a single number.

SCHEIDLINGER, SAUL. *Psychoanalysis and group behavior*. New York: Norton, 1952. Pp. xviii+245. $3.75.

SCHNEIDMAN, EDWIN S., WALTHER, JOEL, & LITTLE, KENNETH B. *Thematic test analysis*. New York: Grune & Stratton, 1951. Pp. xi+320. $8.75.

SMITH, H. B. *The single woman of today: Her problems and adjustment*. New York: Philosophical Library, 1951. Pp. xiv +130. $2.75.

STONE, C. P. (Ed.) *Annual review of psychology*. Vol. 3, 1952. Stanford: Annual Reviews Inc., 1952. Pp. ix+462. $6.00.

TAKALA, M. *Asymmetries of the visual space*. Helsinki: Finnish Academy of Science and Letters, 1951. Pp. 175. 500 mk.

THOMPSON, GEORGE C. *Child psychology: Growth trends in psychological adjustment*. Boston: Houghton Mifflin, 1952. Pp. xxxiii+667. $5.50.

THURSTONE, L. L. (Ed.) *Applications of psychology*. New York: Harper, 1952. Pp. x+209. $3.00.

TINKER, MILES A. *Teaching elementary reading*. New York: Appleton-Century-Crofts, 1952. Pp. ix+351. $3.50.

VAUGHAN, WAYLAND F. *Personal and social adjustment: Foundations of mental health*. New York: Odyssey Press, 1952. Pp. xiv+578. $4.25.

WOLFF, CHARLOTTE. *The hand in psychological diagnosis*. New York: Philosophical Library, 1952. Pp. xv+218. $7.50.

WOLFF, WERNER. *The dream: Mirror of conscience*. New York: Grune & Stratton, 1952. Pp. vi+348. $8.50.

YOUNG, KIMBALL. *Personality and problems of adjustment* (2nd ed.). New York: Appleton-Century-Crofts, 1952. Pp. x+716. $5.00.

Vol. 49, No. 5 September, 1952

Psychological Bulletin

PSYCHOLOGICAL TESTS IN PSYCHOPATHOLOGICAL PROGNOSIS[1]

CHARLES WINDLE
University of Iowa

Prognosis in psychopathology is essentially a prediction specifying a relationship between characteristics of a psychopathological condition and the eventual outcome of the disorder. In the field of psychological tests, the prognosis will relate certain aspects of test performance (such as the number of Rorschach $F+$ responses or the Wechsler-Bellevue IQ) to the degree of improvement in mental health of the patients in question.

Although prognostic information is of great practical as well as theoretical value, investigative work in this field appears discouragingly unorganized. It is hoped that a critical review of that aspect of the problem of most interest to psychologists, viz., the prognostic use of psychological tests, may help investigators plan more fruitful research. Previous reviews (**10, 42, 44, 77, 103, 113, 136, 164**) of this topic have tended to be predominantly descriptive and of limited scope. There remains a need for a comprehensive account of the evidence with sufficiently critical evaluation of it.

Since relatively little work in this field has involved cross validation, evaluation in this review will be based primarily upon comparisons among findings. The material has been organized on the basis of the particular psychological test employed. Although a classification based on either diagnostic category or type of therapy might have been more desirable, such a classification seemed impossible in view of the data available in the studies to be reviewed. Very frequently the patients employed in a single study covered a range of diagnostic categories—either implicitly as when "psychotics" were studied, or explicitly as when two subcategories of schizophrenia were listed and then considered together because no diagnostic differences in prognosis appeared. Furthermore, the meaning of diagnostic labels is often uncertain. A similar

No 4 Part II July 1952 Missing

---

[1] The writer is greatly indebted to Dr. Joseph Zubin for his guidance and encouragement in the preparation of this paper.

confusion exists in respect to type of therapy; results from different therapeutic procedures have often been grouped together and in some cases none has been specified. Consequently, studies have been classified in terms of the psychological tests used. But within each category of psychological test, studies have been classified as far as possible in terms of varieties of mental disorder. Insofar as these diagnostic categories are meaningful, it is possible for the reader to relate prognostic test results to them. As a source of reference and basis for comparisons a master table (Table 1) has been compiled in which some of the most important characteristics of the studies in this field which the author judged to be the best are presented. The text of the article will attempt to elaborate the most important of the specific criticisms implicit in the table and to evaluate the present use of specific psychological tests for prognosis.

The term *prognosis* is used to cover a large range of predictive relationships. The meaning of prognosis is a function of three kinds of variables: (*a*) the situation from which the prediction is made, (*b*) the intervening conditions which may influence the eventual outcome, and (*c*) the final conditions predicted.

Within the first category fall such factors as the nature, duration and severity of the disease, and the patient's age, sex, attitudes, and capacities. Psychological test performance would be included in this category. To test the prognostic efficacy of any one factor it is necessary to hold the others constant, an ideal which usually can only be approximated. Obviously, the more thorough the experimental and statistical control of confounding variables, the greater the value of the study.

Among the more important intervening conditions are the type of therapy employed, if any, and the amount of time elapsing prior to evaluation of outcome. The base line for this evaluation under different therapies is the rate of the so-called "spontaneous remission," or the expected rate of improvement when no specific therapy is applied. It has been reported that regardless of the type of therapy employed, little, if any, enhancement above the rate of spontaneous remissions is gained (**34**, **173**). It seems that at the present state of therapeutic development some characteristic of the individual is the important factor, not the specific therapy. Perhaps if each individual were to be given the type of therapy most suitable to his particular endowment, the rate of improvement might be enhanced. When the prognostic criteria for spontaneous remissions are established, it will be easier to develop the prognostic criteria for specific therapies.

TABLE 1

Summary of Experiments on the Use of Psychological Tests for Clinical Prognosis

When the statistical evaluation provided by the author was unsatisfactory, the significance of the results was calculated by the method of chi square. Asterisks indicate cases in which data were presented but no calculation of significance was made; other symbols indicate cases of apparently inaccurate evaluation in the original articles.

In order to conserve space, abbreviations have been employed in the table. The following key will provide an explanation of these abbreviations.

Abbreviation Key to Summary Table

Type of Patients

A —alcoholics
AC —asthmatic children
CSc —child schizophrenics
DisSc —discharged schizophrenics
Dp —depressed patients
DRPD —delayed recovery from physical disease
DRT —delayed recovery from thyroidectomy
GP —general paretics
H —hysterics

IM —involutional melancholics
MD —manic depressives
MN —mixed neurotics
MtD —mental defectives
N —neurotics
NRT —normal recovery from thyroidectomy
OP —other psychoses
PC —problem children
PH —physically handicapped

PNSc —pseudoneurotic schizophrenics
Pp —psychiatric patients
PrG —promiscuous girls
Ps —psychotics
PsN —psychoneurotics
PT —pulmonary tuberculous
Sc —schizophrenics
S & M —suicide and murderer

Duration and Follow-up Time

CT —concurrent trends
EP —"extensive period"
Im —immediate
L —long

(M) —mean
(Md) —median
Sev —several specific times
Sh —short

SM —several months
V —varied
(σ) —standard deviation

Therapy

AT —allergic treatment
B —brief
BECT —brief stimulus ECT
BS —brain surgery
CF —convalescent furlough
ECT —electric convulsive therapy
En —electronarcosis
I —insulin shock therapy
Ma —malaria treatment
Mz —metrazol shock therapy
No —no specific therapy
Pa —psychoanalytic
PR —physical rehabilitation

Ps & CW—psychiatric & casework service
Py —psychotherapy
S —shock
TP —"total push"
V —varied
WP —wage placement educational program

Prognosis

F —favorable
U —unfavorable

Miscellaneous

NG —information not given

Statistical Significance

D —data
G —given
NG —not given
p —probability
SS —"statistically significant"
* —$p < .01$ by chi square
** —$p < .05$ by chi square
*** —$p < .10$ by chi square
**** —$p > .10$ by chi square

TABLE 1, continued

Experimenter	*N*	*Type of Patients*	*Duration of Illness*	*Follow-up Time*	*Therapy*	*Prognostic Index*	*Prognosis*	*Statistical Significance*
						Rorschach Test		
Piotrowski (123)	18	Sc	NG	NG	I	$\geqq 3\ W+$ responses	F	$p<.02$#
	18	Sc	NG	NG	I	$\geqq 1\ M+$	F	$p<.01$†
Piotrowski (124)	17	Sc	NG	NG	I	$F+\%$ of 70 or higher	F	$p<.10$‡
	60	Sc	NG	NG	I	Discrepancy between actual and potential function, lack of concentration, good percepts with uncontrolled associations, abnormal distribution of color responses	F	p—NG D—NG
Piotrowski (125)	94	Sc	Sh	Im		$\geqq 1$ meaningful color response	F	SS*
						$\geqq 1$ human movement response	F	p—NG D—G***
Piotrowski (126)	60	Sc	NG	SM	I	Moderate deviation from normal; fearful inhibition of responsiveness (blind prognosis)	F	p—NG D—G*
	104	Sc	NG	Sh	I	Variety, differentiation of generic and specific terms, fit, color responses, indirect color approach, demurring	F	p—NG D—G*
						Direct response to black and dark gray colors; human movement	F	p—NG D—NG
Piotrowski & Lewis (128)	100	Sc	Sh	3 yr.	V	High weighted score: [+3 ($hEvd$); +2 ($FC>2$; $W+=3$ or 4, $M=2$, and $\Sigma c \geqq \Sigma C$; $W+=5$ or 6 and $M>5$; $h\%>1<20$); +1 (Ind$P$; $FC=1$ or 2; $(M+\Sigma C)/(M-\Sigma C)>\pm 2$; $W+=3$ or 4 and $M>1$; difference between any two of ΣM, ΣC and $\Sigma c<2$, unless each is 0; both $\Sigma c<\Sigma C$ and $(T/R-\text{Int}RT)>30''$; $a\%<35$); −1 (IntelPas; $M=0$; $h\%=0$)]	F	$p<.01$
Pacella *et al.* (118)	20	Sc	NG	1 mo.	ECT	Least deviation from normal in M, CR and $F+\%$	F	p—NG D—NG
Lourie *et al.* (98)	20	CSc	10 mo. (M)	4–11 yr.	V	Unwilling or unable to cooperate on the Rorschach	U	p—NG D—NG
Kisker (82)	28	Ps	NG	NG	I & Mz	Agreement with Piotrowski (124)		p—NG D—NG
Halpern (61)	17	Sc	Sh	1–3 yr.	I	High productivity (R); emotionality (color sum, c, C', K, M); and empathy (H)	F	p—NG D—G**
Morris (111)	41	Pp	3.9 yr. (Md)	½–1 yr.	Mz	$\geqq 70\%$ well perceived responses ($R+$)	F	$p=.08$‡
						$\geqq 13$ detail responses (D)	F	$p=.07$‡
						$\geqq 1$ space response (S)	F	$p<.06$‡
						$<2\ CF$	F	$p<.05$§
						$<15\%$ anatomy responses	F	$p<.02$#
						$<17\%$ poorly perceived original responses ($O-$)	F	$p<.03$§
						Weighted score based on these six signs		$p<.03$ (?)†
Oppenheim & Brower (117)	35	Ps-Dp	NG	Sh	ECT	Meaningful color responses	F	p—NG D—G***
Van Bark & Baron (166)	NG	Ps	NG	NG	NG	Neurotic elements like color shock	F	p—NG D—NG
Graham (57)	18	Sc	3 yr. (Md)	Im	I & Mz	Many chiaroscuro responses	F	p—NG D—G*
McCall (102)	27	Ps	L	3½ yr.	BS or TP	High *Plimp* (plant importance)	F	$p<.02$
						High P (popularity of response)	U	$p<.05$

† Calculated with Yates' correction: $p<.02$.
# Calculated with Yates' correction: $p<.05$.
§ Calculated with Yates' correction: $p<.10$.
‡ Calculated with Yates' correction: $p>.10$.

TABLE 1, continued

Experimenter	N	Type of Patients	Duration of Illness	Follow-up Time	Therapy	Prognostic Index	Prognosis	Statistical Significance
						Long *CRT* (reaction time to colored cards)	F	$p<.05$
						Long *TRT* (reaction time to all cards)	F	$p<.05$
						High *DeH* (dehumanization)	U	$p<.05$
						Ascendance attributed to perceived figures	F	$p<.05$
						High *SuC* (surface color)	F	$p<.05$
Katz (**80**)	113	Sc	NG	Im	I	No Rorschach prognostic index		
Rabin (**130**)	10	Ps	NG	Im	ECT	Single Rorschach factors not prognostic		
Wechsler *et al.* (**169**)	15	Sc	Sh	Im & 6–18 mo.	I	No Rorschach prognostic index		
Rioch (**134**)	36	Pp	NG	NG	Py	No Rorschach prognostic index		
Rees & Jones (**131**)	35	Sc	NG	Im & $\gtreqless$3 yr.	I	Piotrowski's 1941 criterion ($\geqq$3 of 6 signs) (**126**)	not prognostic	$p<.5>.3$
	19	Sc	NG	Im & $\gtreqless$3 yr.	En	Piotrowski's 1941 criterion ($\geqq$3 of 6 signs) (**126**)	not prognostic	$p<.7>.5$
	22	Sc	NG	Im & $\gtreqless$3 yr.	ECT	Piotrowski's 1941 criterion ($\geqq$3 of 6 signs) (**126**)	not prognostic	$p<.8>.7$
	10	Sc	NG	Im & $\gtreqless$3 yr.	BS	Piotrowski's 1941 criterion ($\geqq$3 of 6 signs) (**126**)	not prognostic	$p<.3>.2$
	86	Sc	NG	Im & $\gtreqless$3 yr.	V	Of >20 signs, none prognostic[1]		p—G
Scherer (**141**)	41	Sc & MD	V	2–4 wks.	BECT	Multiple-choice Rorschach indices	not prognostic	$p>.05$
Scherer (**142**)	26	Pp[2]	NG	2–4 wks.	No	Low rejection	F	$p<.001$
	22	Pp	NG	NG	BS	High *M*	F	$p<.001$
	22	Pp[2]	NG	NG	No	Multiple-choice Rorschach indices	not prognostic	$p>.05$
Filmer-Bennett (**44**)	24	Pp	NG	>1 yr.	V	Of >50 signs only high *R* to cards VIII, IX, & X	F	$p<.05$
	54	Pp	NG	>1 yr.	NG	Five psychodynamic cluster patterns		$p>.20$ (each)
						Seven patterns of signs		$p<.02$ (each)
						Combination of the best of the above patterns of signs		p—NG D—G****
						Individual consideration of 7 pattern variables		p—NG D—G
	78	Pp	NG	>1 yr.	V	Basic Rorschach Score of Buhler, Buhler, & Lefever	not prognostic	$p>.05$
Roberts (**135**)	51	N (?)	NG	Im	Py	Eleven indices derived from the literature	not prognostic (each)	
Siegel (**154**)	96[3]	PsN	NG	Im	PaPy	Of many separate signs only 3 (*FC*, *Aobj*, & *R*) prognostic		SS (each)
						Of several trait syndromes only 3 prognostic		SS (each)
						Five blind predictions	not prognostic (each)	
Dickson (**36**)	35	PsN	NG	Im	Py	High *H*%	F	$p=.02$
						High *P*, presence of *Fc*	F	$p=.05$ (each)
						Presence of *M*, and 3 or more *M*	F	$p=.01$
						$F\%>80$, presence of *CF*	U	$p=.05$ (each)
						$FC>(CF+C)$	F	$p=.10$
Bradway *et al.* (**18**)	16	PrG	NG	EP	Ps & CW	(20 or more *R*; mean response time <40 secs; $FM\geqq2$; *FK*; *Fc*; $P\leqq4$; $D\%\geqq50$; $FC\geqq CF$; no rejected cards; and %*R* to last 3 cards $\geqq40$)	F	p—NG D—G
						8 or more of these 10 signs	F	p—NG D—G*
	20	PrG	NG	EP	Ps & CW	5 or more of these 10 signs	F	p—NG D—G**
Krugman (**90**)	NG	PC	NG	NG	NG	No single factor can determine treatability; 19 factors or aspects listed as prognostic		p—NG D—NG
Siegel (**153**)	26	PC	NG	1 yr.	Py	Refusals, *FC*, *W*%, *Fc*, *H*, *O*, *F*+%, and favorable response to Testing the limits	F	p—NG

---

[1] The authors concluded that "a number of features showed statistically significant differences between the recovered and not improved groups, but none of these showed a statistically significant predictive value" (**131**, p. 689). This conclusion may be subject to disagreement.

[2] Chosen to match the group above.

[3] Subdivided to form original and cross-validating groups; these indices applied to both groups.

TABLE 1, continued

Experimenter	*N*	*Type of Patients*	*Duration of Illness*	*Follow-up Time*	*Therapy*	*Prognostic Index*	*Prognosis*	*Statistical Significance*
Billig & Sullivan (13)[4]	29	A	NG	1 yr.	NG	$O-$, $CF-$, Shading Shock, and CF	U	D—NG
	11			none		High score: $\left[2\left(\frac{W+}{W}\leqq 33\tfrac{1}{3}\%;\ D<50\%;\ Dd+S \geqq 10\%;\ F\geqq 60\%\right);\ 1\left(k>1;\ K>1;\ c>1;\ M<FM;\ FC\leqq 1;\ FC<(CF+C);\ \frac{\text{VIII+IX+X}}{R}<30\%;\ Hd>H;\ \underset{\text{sex}}{p^{1}}>1;\ At>1;\ W+<M\right)\right]$	U	p—NG D—G
Sillman (155)	42	A	NG	2 yr.	NG	Rigidity in responses; blocked efforts after emotional gratification; low capacity for extraversion, verbalization, identification and love; more anxiety, internalized aggression and sense of inferiority	U	p—NG D—NG
Sloan (159)	30	MtD	—	SM	WP	Beck's criteria of good adjustment: ($F+\geqq 50\%$; $P=4$ or 5; no C; $CF=1$ or 2; C sum <3; $A\%$ 50–80; $M<2$)	U	$p<.05$
						Individual quantitative indices	not prognostic	$p>.10$
						Recoverability from color and shading shock	F	p—NG D—G*
Harris & Christiansen (67)	35	DRPD	—	Im	BPy(Pa)	Location, determinant and content scores	not prognostic	
						Empirical weighted score: $[+2(M, FM,$ Anat-sex; $+1.5(FK)$; $+1(Fc, C, C')$; $-.5(F)$; $-1(K, F-, Rej)$; $-2(c, FC)$; $-2.5(CF)]+4$ or more	F	$p<.01$†
Levi (93, 94)	20	PH	NG	NG	PR	Subjective evaluation (movement and no color)	F	$p<.05$
						Many anatomy responses (60% or more)	U	p—NG
Ellis & Brown (39)	26	PT	NG	CT	—	Normal pattern; neurotic pattern with guilt feelings	F	D—NG
						High R, high Z	F	$p=.01$ (each)
						High M, high sum C, high P, many different content categories	F	$p=.05$ (each)
						Rejection of card	U	p—NG D—G*
Rabin (129)	1	S & M	—	2 mo.	—	Absence of emotional expression and evidence of psychic tension (color and shading shock)	U	p—NG D—G
Davidson & Conkey (32)	20	GP	NG	1 yr.	Ma	Piotrowski's signs of organic disorder	U	p—NG D—NG
						Free Drawing Test		
Fiedler & Siegel (43)	34	PsN	NG	NG	Py	Goodenough's criteria of drawing the head are met	F	$p=.03$
						Minnesota Multiphasic Personality Inventory		
Harris (64)	NG	Pp	NG	NG	S	High scores in psychasthenia and psychotic scales	U	p—NG D—NG
Harris *et al.* (66)	31	Pp	L	Im	En	High scores on only D scale or on several scales including D, Sc, Pt, and Pa	F	p—NG D—NG
						Normal scores on all scales	U	p—NG
						Sorting based on criteria "developed in other studies"		D—G**

---

[4] The prognostic indices listed by Billig

another report (12) of the

TABLE 1, continued

Experimenter	*N*	*Type of Patients*	*Duration of Illness*	*Follow-up Time*	*Therapy*	*Prognostic Index*	*Prognosis*	*Statistical Significance*
Hales & Simon (60)	20	Sc	3 yr. 2 mo. (M)	Im	I	MMPI scores and profiles not prognostic High scores in psychotic components	F	p—NG D—NG
Pearson (119)	29	Sc	NG	Im	ECT	Score on *D* scale >84 or <55; Score on *Sc* scale >84; Scores ≥80 on 4 or more scales	U	$p<.01$ (?)∮
	15	Sc	NG	Im	ECT	Validation of the above empirical finding		p—NG D—G***
Pearson & Swenson (120)	33	Pp	NG	Im	ECT	Validation of the above empirical finding		$p<.01$
Feldman (41)	40	Pp	NG	1 yr.	S	42 MMPI items differential of outcome		$p=.05$ (each)
	44	Pp	NG	NG	NG	Validation of this scale		$p<.01$
Feldman (42)	84	Pp	NG	3, 6, & 12 mo.	S	*Ps* scale (52 MMPI items) differential of outcome		$p<.01$
						High *Pd*, *Pt*, *Sc*, & *Ma* scores; unfavorable in *F* & *K*	U	SS (each)
	100	Pp	NG	6 mo.	S	Validation of *Ps* scale		$p=.0002$
	53	PsN	NG	Im	BPy	Abbreviated *Ps* scale differential of outcome		$p=.02$–$.01$
	100	Pp	NG	6 mo.	S	Blind sorting of MMPI records for prognosis		$p=.001$
Pacella *et al.* (118)	40	Sc	NG	1 mo.	ECT	Pre therapy MMPI Prepost drop of 10 points on *Hs*, *D*, *Hy*, & *Pd*	not prognostic F	 p—NG D—NG
Carp (20)	37	Sc	13.7 mo. (Md) 11.4 mo. (σ)	NG	I	Low *Pa*, *Pt*, and *Sc* scores; high *K* scores; low (*Sc*+*K*) scores	U	$p<.05$ (each)
Schofield (146)	48	PsN	NG	NG	NG	Therapy-susceptible items (*Px* scale)	not prognostic	
Harris & Christiansen (67)	53	DRPD	—	Im	BPy (Pa)	High scores on *Pd*, *PdR*, *Pa*, *Sc*, & *Ma* scales	U	$p<.05$ (each)
						Blind classification by clinical criteria		$p=.20$ or $<.02$
						8 of 35 empirically derived subscales		$p<.05$ (each)
Ruesch *et al.* (138)	20 18	NRT DRT	—	NG	—	Delayed recovery cases have high scores on *Pd*, *Pt*, and neurotic scales		$p<.02$
						Measures of Aggression		
Albee (3)	127	Pp	NG	>1 yr.	NG	Extrapunitive behavior	U	$p<.001$
	81	Sc	NG	>1 yr.	NG	Extrapunitive behavior	U	$p<.001$
Albee & Goldman (5)	44	Pp	NG	NG	NG	Picture-Frustration E% and I%	not prognostic	
						Ability Tests		
Langfeldt (92)	83	Sc	NG	7 yr.	NG	Ratings of "clever" or "intellectually debilitated"	U	p—NG D—G**
Terry & Rennie (163)	47	Sc	NG	4–5 yr.	NG	Intelligence	not prognostic	
Malamud & Render (105)	177	Sc	NG	5–9.5 yr.	NG	Binet IQ above 105	F	p—NG D—G*,**
	309			V		Binet "scatter" of 5 years or more	U	p—NG D—G****
Graham (57)	9	Sc	3 yr. (Md)	$\frac{1}{3}$–3 mo.	I & Mz	Malamud-Palmer index on Binet; Binet scatter; vocabulary test; digits forward & backward; free association; arithmetic; interpretation of fables Detection of verbal absurdities	not prognostic F	 p—NG D—G**
O'Connell & Penrose (116)	30	Pp	NG	NG	Mz	Incompetence in reaction time to auditory stimuli, in tapping rate and in strength of grip	F	p—NG D—NG
Schnack *et al.* (143)	31	Sc	NG	V	I or Mz	A set of Binet, Word Association and aspiration test signs		p—NG D—G
	23–50	Sc	6.1 mo. (M) 16.5 (σ)	Sev	I or I+Mz	The above set of signs High scores in 4 of 13 signs from the above tests	not prognostic F	 $p<.10$

TABLE 1, continued

Experimenter	*N*	*Type of Patients*	*Duration of Illness*	*Follow-up Time*	*Therapy*	*Prognostic Index*	*Prognosis*	*Statistical Significance*
	14–22	Sc	7.2 mo. (M) 16.5 (σ)	Sev	Mz	More insulin (the above 13) than metrazol (the below 11) signs Low scores in 6 of 11 different signs	F F	$p = .05$ $p < .10$
Lourie *et al.* (98)	20	CSc	10 mo. (M)	4–11 yr.	V	More matrazol than insulin signs Stanford-Binet <70 or child mute	F U	$p = .20$ p—NG D—G*
Kriegman & Hilgard (89)	30	PC	NG	≧7 mo.	Py	Higher performance scores (Merrill-Palmer, Arthur, and Pintner-Patterson) than verbal (Stanford-Binet)	F	p—NG D—G***
Rabin (130)	10	Ps	NG	NG	ECT	Stanford-Binet IQ≧110	F	$p < .01$
Peters (121)	71	Ps	L	>1 yr.	BS	Wechsler-Bellevue IQ and memory scales Ratings of normal intelligence	not prognostic not prognostic	
Harris *et al.* (66)	24	Pp	L	Im	En	Abbreviated Wechsler-Bellevue IQ Higher scores on untimed than timed tests	not prognostic F	 p—NG D—G***
Malamud *et al.* (106)	47	IM	1 yr.	SM	Mz or Py	Estimated intelligence dull normal or lower	F	p—NG D—G***
Malamud *et al.* (107)	44	IM	1 yr.	7 yr.	Mz or Py	Estimated intelligence below average	F	p—NG D—G***
Fisher (45)	16 43–8	DpPp DpPp	NG NG	Im Im	ECT ECT	Attitude of feeling bad (one of two 'scales used) Wechsler-Bellevue IQ, Word Naming, Counting by 3's, Word Association, and Level of Aspiration	F not prognostic	$p < .05$
Scherer (141, 142)	38–41	Sc & MD	V	2–4 wk.	BECT	Of >17 tests only lower scores on delayed than immediate memory Efficiency on Minnesota Rate of Manipulation Test Efficiency on Capps' Categorization Test	U F F	$p = .05$ $p = .01$ $p = .01$
Willenson (172)	21–25 21–22 22 110	Pp^2 Pp Pp^2 Pp	NG NG NG NG	2–4 wk. NG NG NG	No BS No BS or TP	Of >17 tests only high visual memory Of >17 tests ten tests of memory and abstraction prognostic Of >17 tests High Wechsler-Bellevue IQ	F none prognostic F	$p < .05$ $p < .05$ (each) $p > .05$ (each) $p < .01$
Carp (21)	9 42	PNSc Sc	NG 19.2 mo. (M) 12.8 mo. (σ)	NG 3 mo.	BS I	High Similarities, Picture Arrangement, Block Design, and Object Assembly scores Wechsler-Bellevue IQ High Wechsler-Bellevue IQ	F not prognostic F	$p < .05$ (each) $p < .01$
Piotrowski & Lewis (128)	91	Sc	Sh	3 yr.	V	IQ below 90	U	p—NG D—G****
Zubin & Windle (180)	16	Ps	L	2 yr.	BS or TP	Good performance on Metenym Test	U	$p < .05$
Landis & Erlick (91)	35 16	Ps Ps	L L	2 yr. 2 yr.	BS or TP BS or TP	Weighted score derived from criterion groups "Confusion errors" in the Porteus Maze Test	prognostic F	$p < .05$ p—NG D—NG
Windle & Hamwi (173)	28 26	Ps Ps	>1 yr. <1 yr.	1 & 3½ yr. 0–9 mo.	BS or TP Py or S	Complex Reaction Time Test score >30 Complex Reaction Time Test score >30 Divergence between these two trends	U F significant	$p < .05$ $p < .05$ $p < .05$
Malamud & Gottlieb (104)	80	H	NG	NG	Py	Binet IQ above 90	F	$p < .10$
Klugman (86)	64 104	MN PsN	NG NG	NG Im	Py CF	Binet IQ above 90 High Army General Classification Test score	F F	$p < .20$ p—NG D—NG
Miles *et al.* (110)	33	N	> 5 yr. (M)	>2 yr.	Py	High Wechsler-Bellevue IQ	F	$p < .01$

Chosen to match the group above.

TABLE 1, *continued*

Experimenter	*N*	*Type of Patients*	*Duration of Illness*	*Follow-up Time*	*Therapy*	*Prognostic Index*	*Prognosis*	*Statistical Significance*
Dickson (36)	67	PsN	NG	Im	Py	IQ	not prognostic	
Ruesch *et al.* (138)	23	NRT	—	NG	—	High IQ	F	$p<.05$
	20	DRT	—	NG	—			
Harris & Christiansen (67)	53	DRPD	—	Im	BPy (Pa)	Abbreviated Wechsler-Bellevue IQ	not prognostic	
Hamlin (62)	101	MtD	—	NG	NG	High $\frac{\text{mental age on non-verbal tests}}{\text{Binet MA}}\times 100$	F	p—NG D—NG
	48	MtD	—	NG	NG	The above pattern score $>105<124$ The above pattern score <80	F U	p—NG D—G*
Earl (38)	66	MtD	—	NG	—	Psychographs on 2 verbal and 2 performance tests: unbiased or biased in favor of performance tests irregular or biased in favor of verbal tests	 F U	 p—NG D—G*
					Abstraction Tests			
Bolles *et al.* (16)	19	Sc	27 mo. (Md)	Im	I	Good performance on the Vigotski, Weigl and BRL sorting tests	F	p—NG D—G*
Graham (57)	7	Sc	3 yr. (Md)	$\frac{1}{3}$–3 mo.	I & Mz	Bolles battery	not prognostic	
Zubin & Thompson (179)	49	Sc	Sh	Im & 6 mo.	I	Good performance on Bolles battery	F	SS
	43	Sc	L	Im & 6 mo.	Mz or I+Mz	Bolles battery	not prognostic	
					Miscellaneous Tests			
Low *et al.* (99)	102	Sc	NG	4.5 yr.	NG	Appropriate verbal response to Life Situation Test	F	p—NG D—G*
	54	MD	NG	4.5 yr.	NG	Appropriate verbal response to Life Situation Test	F	p—NG D—G****
Adler (2)	47	OP	NG	4.5 yr.	NG	Life Situation Test	not prognostic	
	50	Sc	NG	—	NG	Functional matching of pictures correlated with good prognosis ratings		p—NG D—NG
Cattell (23)	64	PC	NG	>1 yr.	NG	Low score on perseveration tests	F	p—NG D—G*
	57	PC	NG	>1 yr.	NG	Extreme scores in fluency tests	F	p—NG D—G**
	24	PC	NG	>1 yr.	NG	For "nervous children," low "F" scores	F	p—NG D—G***
Voth (167)	209	Sc	NG	NG	Mz or No	Extensive or no autokinetic movement Many directional changes	U U	$p=.01$ $p<.02$
	259	MD	NG	NG	Mz or No	Extensive or no autokinetic movement	U	$p=.02$–$.05$
	43	GP	NG	NG	NG	No autokinetic movement	U	$p=.05$–$.10$
	71	N	NG	NG	NG	Autokinetic movement	not prognostic	$p=.30$–$.50$
	754	Pp	NG	NG	NG	Extensive or no autokinetic movement	U	$p<.01$
Balinsky (8)	21	DisSc	11 mo. (Md)	1.5 yr. (M)	—	Positive attitude toward status in society, high drive, emotional expressiveness, small discrepancy between vocabulary & dexterity-spatial tests, and undepressed motility	F	p—NG D—G
Peters (121)	71	Ps	L	>1 yr.	BS	Ratings of + for mood swings and hallucinations; — for energy, purposive, active, initiative, evident interests, and reality grasp	U	$p<.05$ (each)
						Ratings of + for restless, aggressive, agitation, and speed or flight of ideas	F	$p<.05$ (each)
Harris *et al.* (63)	14	AC	1–9 yr.	NG	AT	Crying easily and confidence in mothers	not prognostic	

Estimates of final outcome of mental diseases may differ when different criteria of improvement are adopted. An individual rated as improved on the basis of socioeconomic status may be considered unimproved on the basis of psychosomatic complaints. In some cases emphasis is placed on the objectivity or reliability of the criteria and indices such as the hospitalized status of the patient or a Distress-Relief Quotient are used. More frequently the emphasis on validity leads to the rejection of such mediated criteria, and clinical opinion serves as the accepted index. Clinical opinion, however, fails to lend itself to accurate description or reproduction, and thus is of dubious reliability among investigations. As yet the problem of optimal measures of adjustment remains unsolved. It is necessary to bear in mind that unless measures of outcome are highly correlated, the meanings of prognoses in different studies will differ.

RORSCHACH INK BLOT TEST

The psychological test which has been used most often in the search for prognostic indices is the Rorschach. According to Hertz, "It is in its prognostic power, perhaps, that the Rorschach has its greatest possibilities"(**70**, p. 677). If Rorschach signs such as the $F+$ score can give us information about ego strength, emotional control, and other personality factors psychiatrically said to be important for recovery (**9**, **69**), the Rorschach test warrants the large amount of attention it has received. It is well to remember that the reason for the many claims for the Rorschach may not be solely the sensitivity of the test to important personality variables. Some influence can be attributed to the freedom of interpretation permitted, the frequent immunity of the results to statistical treatment, and the large number of discrete indices whose sheer number is bound to produce statistical significances if the .05 level of confidence is accepted.

Psychoses. The leading claim for the prognostic efficacy of the Rorschach technique with psychotics is that voiced by Piotrowski in a series of eight articles (**98**, **118**, **123**, **124**, **125**, **126**, **127**, **128**). Three descriptions of the crucial personality differences between outcome groups have been propounded: (*a*) one based on functioning at or below capacity (**124**), (*b*) the second based upon a distinction between "emotional" and "intellectual regression" (**126**), and (*c*) the third based on a change in either the "variable" or "constant" personality traits (**127**). Evidence of subnormal functioning at the patient's optimal level of capacity, intellectual regression, and change in the personality traits normally expected to remain constant were held to be unfavorable for improvement. This prognosis was thought to apply regardless of type of therapy (**118**).

Apparently Piotrowski tried to employ an objective method of revealing those personality characteristics related to outcome. He did this by stipulating retrospectively those quantitative Rorschach signs which showed predictive value. The logic and the substance of this application of experimental procedure are questionable. Aside from such methodological flaws as the omission of Yates's correction when using the chi-square test with small samples (**123**), there is an almost complete neglect of cross validation of empirical signs. All but one of this group of studies are ex post facto in nature, the relationship between predictive test and outcome having been determined after both were known. When this is the case, the experiment must be repeated to determine whether statistically significant results may not be due to sampling error. This need for cross validation becomes clear in view of the considerable fluctuation among the empirical indices reported by Piotrowski in his several studies.

The one attempt at a confirmatory experiment consisted of "blind" prognoses which were reported to be based upon information gained from earlier experiments (**126**). The Rorschach criteria employed to predict improvement were: "(A) a moderate deviation of personality from the norm of healthy adults, and (B) inhibition of responsiveness to the environment caused by the schizophrenic's fear that the disease process will prevent him from maintaining his former adjustments" (**126**, p. 808). These criteria, however, do not appear to conform very closely to Piotrowski's previous generalizations nor do they conform to the previous empirical signs (see Table 1). In view of this divergence, it is doubtful whether this study can be considered confirmatory. It is difficult to understand why Piotrowski failed to utilize the previously found empirical indices or how the nebulous criteria he preferred could be objectively handled.

Piotrowski's latest study (**128**) appears to constitute still another revision of his position. Fifteen signs, most of them new in structure and/or content, were weighted to form a battery which was, again retrospectively, differential in respect to outcome. This new battery remains to be validated.

Numerous authors have expressed support for Piotrowski's descriptions of the personality traits related to prognosis. Their studies, however, have either lacked evidence for their conclusions (**11**, **82**, **117**, **168**) or employed different, if not opposing, signs from those Piotrowski used (**61**, **111**, **164**). Only three of the studies supporting Piotrowski's views (including those done by Piotrowski) stated the statistical significance of the findings, and two of these (**111**, **123**) disregarded Yates's correction for small entries in chi-square tests (see Table 1).

In contrast, there are two studies in which the authors subordinated interpretation to results. Thus, Graham (**57**) found that chiaroscuro responses indicated favorable prognosis but preferred not to identify this index with any hypothesized personality variable, although

Piotrowski (**126**) felt chiaroscuro was implied in his signs and therefore in his theory. McCall (**102**) departed even more radically from traditional findings and interpretations. Using rating scales of various personality and perceptual dimensions, he found that the few prognostic indicators which did appear pertained to various aspects of the *content* of Rorschach responses which receives relatively little emphasis in orthodox methods of scoring and interpretation.

In even greater contrast, some investigators have reported an absence of prognostic value for the Rorschach technique (**44, 80, 130, 131, 135, 136, 143, 171**). The most impressive of these studies (**44, 131**) consist of thorough attempts to verify the prognostic efficacy of the signs reported by Piotrowski and others. The fact that these rigorous confirmatory studies tend to be negative militates against the acceptance of the rather unconvincing positive conclusions which have been advanced.

Neuroses. Of the five articles found reporting prognostic efficacy of the Rorschach specifically for neurotics, two were no more than hints for future research (**46, 85**). Another (**43**) was nullified when cross validation proved the previously reported sign to have no validity.[2] A report by Dickson (**36**) lists six signs found to differentiate significantly between outcome groups. Siegel (**156**), on the other hand, found that a very low percentage of signs showed prognostic value in both an exploratory and cross-validating study. Those signs that did have general predictive power differed from those of Dickson. The relationships between "trait-syndromes judged from the entire protocol" and improvement in psychotherapy were reported to be slight, although statistically significant. None of the blind predictions of improvement was significant.

Roberts (**136**) has tried to verify previous prognostic claims for the Rorschach in a homogeneous group of patients who seem best to fit the category of neurotics. None of the eleven indices derived from the literature proved prognostic. This might have been expected since several of Roberts' hypotheses originated in studies specifically dealing with psychotics.

Perhaps related to a discussion of neuroses is a study of the treatability of personality difficulties related to sexual promiscuity (**18**). Rorschach signs relevant to treatability, generally similar to those found by Dickson, were derived from criterion outcome groups, the membership of each group being restricted to women who were judged to show "intrapsychic conflict." These criteria maintained their prognostic power when applied to another sample of promiscuous women.

Problem children. None of the six Rorschach reports on the prognosis of problem children can be considered satisfactory. Four of the articles (**90, 152, 153, 154**) read more like advertisements than experiments. The fifth (**155**) presents evidence for the alleged prognostic power of twelve Rorschach factors in such form that it cannot be evaluated, and the sixth (**141**) does not specify the Rorschach signs used.

---

[2] SIEGEL, S. M. Personal communication, 1950.

Alcholism. There have been two studies concerned with the prognostic use of the Rorschach for alcholics. These are in agreement that "extraversion" is a favorable sign and "rigidity" unfavorable. The agreement between findings is clear, however, only on the level of interpretation. The article by Sillman (**157**) provided no evidence for the prognostic claims and did not specify the experimental indices basic to the interpretations. It also appears that the findings were not derived wholly from the Rorschach record. The articles by Billig and Sullivan (**12, 13**), on the other hand, are replete with data. But there is some inconsistency with respect to the prognostic indices between the two articles (**12**, p. 126; **13**, p. 572), and statistical evaluation of either the individual indices or the total battery is lacking.

Apparently all that can be said is that we have a few hints as to what personality factors may be important for prognosis in alcoholism, the most likely being "extraversion" and "rigidity." In concurrence with these hints, Hoch (**73**) has reported that in 200 cases of alcoholic psychosis extraverts were three times as frequent as introverts and had a prognosis three times as favorable. In this study no attempt seems to have been made to define and measure extraversion objectively. Other suggestions of prognostic personality factors for alcoholics have been even less objective and either not subjected to confirmatory study (**88, 109, 115**) or not confirmed when reinvestigated (**52, 166**).

Somatic illness. The three studies of the prognostic use of the Rorschach for somatic illness are based on the assumption that there is a close relationship between the patient's mental attitude and the ensuing progress of the physical disease. It may or may not be logical to assume that the same mental attitudes lead to improvement in different somatic conditions. In any case, the suggestions provided for each of the different types of illness differ irreconcilably. Thus, Harris and Christiansen (**67**)[3] found among cases showing delay in recovery from physical disease, operation, or accident that poor outcome in psychotherapy was directly related to responses based on the perception of color (color unfavorable), while Ellis and Brown (**39**) found declining tuberculosis patients had less emotional contact with the environment (low sum *C*) than did recovering patients (color favorable). Levi (**93, 94**) claimed that the non-rehabilitated physically handicapped differed from the rehabilitated in showing a high percentage of anatomy responses; but Harris and Christiansen (**67**) reported anatomy-sex responses favorable for psychotherapeutic recovery.

---

[3] It was felt that the Harris and Christiansen study of psychotherapeutic outcome may be fairly compared with those of the course of physical disease since the modification of physical symptoms was among the criteria of improvement.

On the other hand, there is some consistency among the reports. High *M* scores were found favorable and rejection of the stimulus material unfavorable by both Harris and Christiansen (**67**) and Ellis and Brown (**39**). Here also confirmatory studies are desirable.

Mental deficiency. The one Rorschach study of the prognosis of mental defectives presents another striking example of how inadequate the Rorschach test appears to be when subjected to predictive checks. Using employed mental defectives, Sloan (**161**) tested seven criteria (see Table 1) suggested by Beck as prognostic of good adjustment. Beck's criteria significantly differentiated the two outcome groups, but in the unhypothesized direction. Patients who were successful in staying out on wage placement had a reliably greater number of *disagreements* with Beck's criteria than did those who were returned as failures. Sloan felt that "such an absurdly contradictory finding must be attributed to inadequacy of sufficient differentiation in the two groups and, therefore, should not be accepted at face value" (**161**, p. 306). It is difficult to understand why Sloan felt justified in discounting this finding while advocating as an important prognostic clue his qualitative index of color and shading shock based on the same inadequately differentiated groups.

General paresis. The Rorschach technique has also been said (**32**) to be of prognostic value for general paresis. Although prognostic criteria have been outlined, no definitive study has been undertaken to demonstrate this relationship.

Suicide. There have been several Rorschach studies of suicidal tendencies, the most thorough being those of Hertz (**71, 72**), but these will not be discussed here since they are more often diagnostic than prognostic. In order to orient the study of suicide around prognosis, it will be necessary to obtain more cases such as the one reported by Rabin (**129**) in which a Rorschach syndrome combining color and shading shock *preceded* murder and attempted suicide. It also seems necessary to differentiate more carefully between serious and non-serious attempts at suicide (**40**).

Summary. The foregoing review of the prognostic utility of the Rorschach has failed to disclose any very encouraging concordance among studies for any diagnostic category. A considerable number of the positive claims cited in the literature appear to be due to an uncritical attitude concerning validity. Many others may easily be due primarily to chance. When both of these variables are controlled, very few positive results remain.

FREE DRAWING TEST

Fiedler and Siegel (**43**) have presented evidence that psychoneurotics who draw the face and head of a figure "primitively" have a relatively poor chance for improvement. The authors interpreted poor performance in drawing the face as "indicative of inability to form that interpersonal relationship between patient and therapist which is the necessary context of the therapeutic process." Before such an interpretation can be accepted, this study should be repeated and more data relating drawing performance to personality assembled.

Minnesota Multiphasic Personality Inventory

Although the MMPI is somewhat less subject than is the Rorschach to the fluctuations brought about by an infinity of ready-made indices and interpretations, there is enough freedom to permit considerably more exploration than confirmation and more disagreement than agreement. This fact can be seen from the results summarized in Table 1. It is apparent that there is much conflict among the MMPI indices claimed to be prognostic of favorable outcome in different studies.

Psychoses. High scores on the most frequently cited scale (*Sc*) have been held about equally often to be associated with improvement (**20, 60, 66**) and with unimprovement (**42, 64, 119, 120**). Even more striking is the apparent reversal of attitude on the part of Harris regarding the signs for unimprovement. Whereas initially he (**64**) felt that poor prognosis was associated with a pattern of high scores in the psychotic scales, including psychasthenia, he and others (**66**) later reported that among chronically ill patients the unimproved averaged within the normal range before treatment on all the MMPI scales, in contrast to the improved who had very high scores either on several of the scales or only on the depression scale. Unfortunately, the data in support of either of these conclusions are not available nor was the significance of either trend reported. Although data on the accuracy of blind prognostic sorting have been presented (**66**), the criteria on which it was based are unclear, a situation parallel to Piotrowski's blind sorting (**126**).

In light of evidence from other studies (**175, 183**) on the importance of chronicity in prognosis and of additional information concerning the different types of patients employed in the studies by Harris and Feldman,[4] it was thought that much of the apparent disagreement among the MMPI reports might be attributed to radical differences among the patients in chronicity. This interpretation seems plausible, since at least two of the three studies that reported high scores as generally favorable employed patients of long duration of illness, while at least two of the four studies reporting high scores as generally unfavorable dealt with patients of short duration of illness and/or eliminated patients showing false normal scores. There remains, however, some question of the validity of this explanation of divergent findings. Unfortunately, many of the studies have failed to provide sufficient information in regard to both the conditions of the experiment and statistical evaluation of the results. Further, some findings seem to be inconsistent even when chronicity is considered. Carp (**20**) studied a group of patients of apparently wide range of chronicity (see Table **1**). One would expect this group to be differentiated in a manner similar to acutely ill patients (high scores unfavorable), since most of the recovered cases should have been acutely ill. Harris' (**66**) sorting based on criteria "developed in other studies" (presumably studies dealing with acutely ill patients) would not be expected to be so accurate with chronically ill patients if the hypothesized reversal in prognostic criteria with difference in chronicity applies (see Table **1**).

---

[4] Harris, R. E. Personal communication, 1951.

It is highly questionable whether one is justified in concluding that it has been demonstrated that the scores of any of the regular MMPI scales can prognosticate the outcome of therapy for psychotics (**65**). Probably any such generalization must take into account the role of chronicity (among other variables).

Among the most thorough investigations of the prognostic use of the MMPI are those of Feldman (**41**, **42**) and Pearson (**119**, **120**); both series of studies involved cross validation (see Table **1**). Feldman empirically developed a new MMPI scale to be used for prognosis independent of diagnosis. This scale has been applied to more subjects than are customarily used in prognostic experiments and has proven effective, although some of the efficiency may be due to the combination of unlike diagnostic categories. Further work in prognosis should employ this scale or, just as desirable, previously gathered data should be reanalyzed with it. Pearson was concerned only with immediate response of schizophrenics to ECT, regardless of final outcome. Subsequent study of the prognostic efficacy of the MMPI signs for later outcome should help to establish the relationship between different criteria of outcome.

Neuroses. Two scales have been applied to neurotics to determine prognostic utility (**42**, **148**). Only the abbreviated *Ps* scale of Feldman (**42**) seems effective, but this scale lacks cross validation.

Somatic illness. A number of investigators (**67**, **137**, **138**, **139**) have reported that cases of delayed recovery from physical disease, operation, or accident are characterized by abnormally high MMPI scores, especially on the neurotic scales. The clinical impression of these patients seems to be that they are neurotic. Ruesch *et al*. (**139**) further described the delayed recovery group as containing the hysterical or anxiety type (women) and the dependent type (men).

Harris and Christiansen (**67**), going one step further, found high MMPI scores prognostically unfavorable in psychotherapeutically treated cases of delayed recovery. In order to identify those attitudes contributing to high scores, items were grouped into 35 subscales, eight of which showed significant differences between prognostic groups. The attitudes represented by these eight differentiating subscales were thought to be paranoid and psychopathic, indicating that there may be subclinical psychotic or psychopathic trends in the delayed recovery group with poor prognosis. These findings at present lack cross validation and may have been superseded in importance by Feldman's (**42**) use of the *Ps* scale with the same group of subjects (see above).

Summary. Prognostic studies of psychotics using the MMPI exhibit a large amount of disagreement among conclusions. Evidence indicates that some of this disagreement can be attributed to differences in the type of patients studied. Investigation directly on this question seems essential to establishing a consistent and comprehensive picture of prognosis.

The prognostic MMPI studies of somatic illness seem essentially in agreement and promise considerably greater utility than do the Rorschach studies in this field.

THEMATIC APPERCEPTION TEST

So far as the reviewer has been able to discover, the TAT has not been employed to determine prognostic criteria of remission, even though there has been an attempt to demonstrate with case studies how the TAT could be of prognostic value (**81**). Hartman (**68**) determined biserial correlations between the psychiatric rating of good behavior prognosis and 56 TAT categories for 35 delinquent boys. He found eight "significant" ($p < .06$) correlations. However, a third of these could be expected by chance, and correlations between ratings of prognosis from blind analysis of the TAT and either the experimental or psychiatric rating were .15 or lower.

Masserman and Balken (**108**) reported an "occasional significance" of TAT phantasies for prognosis. They felt that the type of change in phantasy with psychiatric interviews could show the abatement of conflicts and the development of insight. Conversely, poor prognosis would result from untouched intense intrapsychic conflicts or unconscious resistance as expressed in recurring phantasies about sick people who cannot be cured. All in all, it does not appear that objective criteria have been found through which the TAT can be of prognostic use, even though indications that it may be useful are available.

MOSAIC TEST

The Mosaic Test has also been asserted to be of prognostic utility, but again there are no studies that have demonstrated its value in this area. Diamond and Schmale (**35**) have described a technique that is an offshoot of the Mosaic Test to detect the "psychological color blindness" of the schizophrenic. They felt this approach to the thinking disorder of schizophrenics would lead to more accurate prognostic standards than those afforded by sorting tests, but as yet this relationship remains undemonstrated.

MEASURES OF AGGRESSION

Recently two studies have reported prognostic value for measures of aggression. Extrapunitiveness and intropunitiveness measured in terms of acts of aggression or "accident" directed toward others or toward oneself were found to be significantly related to outcome for psychiatric patients and for schizophrenics alone (3). Extrapunitive behavior was unfavorable, intropunitive behavior favorable. Scores of the direction of aggression derived from the Picture Frustration Study were not found to be prognostic, although it was suggested that an antithetical pattern between the two measures (overt and P-F) may point to a poor prognosis, especially when the overt measure is extrapunitive (5). These clues are interesting but require verification and more refined methods of measurement than those employed.

TESTS OF MENTAL ABILITY

Psychoses. Attempts to determine the role of intellectual capacity in clinical prognosis have been more objective than investigations of other phases of personality, mainly because the accepted tests of intelligence are relatively standardized in scoring and interpretation. On the other hand, the extent to which different tests measure common aspects of ability is unclear. There is also a large amount of disagreement in the general findings reported. Often it is claimed that psychotics manifesting the greater intellectual power are more likely to improve (16, 21, 57, 89, 92, 98, 105, 143, 144, 145, 174, 175, 181), but just as frequently no relationship is found (45, 57, 66, 121, 130, 143, 144, 165, 181, 183) or else those who perform less efficiently on ability tests show the better outcome (91, 92, 106, 107, 116, 133, 145, 175, 182, 183).

Evaluation of these reports must take into consideration such procedural flaws as the lack of statistically significant evidence (91, 106, 107, 116, 128, 133, 145). Another factor to be considered is the particular tests employed, or the supposedly different abilities involved. Unfortunately, it is hard to deal with this factor in view of our inadequate knowledge of the structure of general or specific abilities in the mentally ill. This inadequacy is not remedied by the various interpretations accompanying prognostic claims.

If it is assumed that the abilities measured in different studies, despite differences in instruments and techniques, are similar, the contradictions in results are even more disturbing than those found for less objective tests. When one uses the IQ as a prognostic indicator, the index is fairly well defined, and ambiguity in the results cannot easily be attributed to vagaries in interpretation.

Although there is evidence that some of the obtained differences in the prognostic value of intelligence tests may relate to diagnostic differences (**106**, **107**, **174**), to the use of different types of therapy (**144**, **145**, **181**), and to different follow-up periods (**145**, **175**, **183**), these factors are not sufficient to eliminate all of the existing inconsistencies. Probably the most hopeful attack on this problem has been the attempt to fractionate groups of patients into like-minded or like-structured subgroups, each of which is characterized by different prognostic indices. Prognostic studies carried out in the first Columbia Greystone Project indicated an association of *poor* test performance with favorable outcome (**91**, **182**, **183**). It was thought that the chronic status of these patients might be a crucial factor in explaining the inverse prognostic relationship. Consequently, the prognostic implications of the Complex Reaction Time Test for the chronically ill (long duration) Columbia Greystone patients were compared with those for acutely ill (short duration) psychotics (**175**). The fact that chronic psychotics who performed poorly and acutely ill psychotics who performed well later improved offers a possible basis for resolving the apparent contradictions in the literature. This concept would also help to explain the differing prognoses for metrazol and insulin shock therapies (**145**, **181**), since the more chronically ill are usually given metrazol. It may be that temporal duration of illness is not a very satisfactory measure of chronicity, since it is difficult to define or measure and since the disease process probably progresses at different rates in different patients. This measure has yielded some degree of success, however, and is a lead worth following up. It should be noted that the above study has not been cross validated, so that the role of chronicity is not definitely established.

Neuroses. Three studies (**86**, **104**, **110**) have agreed that high intelligence leads to a favorable outcome in neuroses, but despite this agreement the evidence from these studies is not strong enough to justify the conclusion that a relationship has been demonstrated. Klugman (**86**) provided no statistical measure of significance and Malamud and Gottlieb (**104**) reported their findings significant to a lower degree than is usually accepted. In addition, Dickson (**36**) reports the IQ to be of no prognostic value for neurotics.

Somatic illness. The two studies relating intelligence to outcome of somatic illness differ concerning the prognostic value of the IQ. Harris and Christiansen (**67**) found no relationship, while Ruesch *et al.* (**139**) reported that high intelligence is significantly favorable. We lack sufficient evidence to determine whether measures of intelligence can be used prognostically in this connection.

PATTERNS OF ABILITY

The fact that interrelationships among particular abilities or traits may be correlated with tendencies toward adjustment is the rationale for the use of pattern analyses. This approach has been used extensively with the Rorschach technique, greatly increasing the number of apparent differences between groups in exploratory studies. The use of pattern analysis with ability tests provides sorely needed information concerning the relationships of abilities without being dangerously subject to chance "findings" since relatively few combinations exist.

The study of patterns of abilities, although not widely used, has focused largely on patients whose primary difficulty is the loss or lack of ability—the deteriorated and the mentally defective. Two studies (**38**, **62**) of mental defectives have indicated that those whose "performance" scores exceed or equal their "verbal" scores have better prognoses of social adaptability than do those whose verbal scores exceed the performance scores. The same prognostic relationship has been claimed for schizophrenics (**8**, **98**).

Other patterns of abilities investigated for prognostic value include relationships between timed and untimed performance tests (**66**), delayed and immediate memory (**143**, **144**,) and Stanford-Binet subtests (scatter) (**57**, **105**). In no case is there confirmation of the various claims.

MISCELLANEOUS TECHNIQUES

There have been many other suggestions of psychological criteria for prognosis, most of which do not fall conveniently into the previously mentioned categories of tests. Many of these criteria are not objective, but nonetheless merit consideration.

Recently Voth followed up a suggestion by Sexton (**150**) that visual autokinesis may be of prognostic value. Voth (**169**) found that intermediate degrees of apparent movement were favorable for psychotics. Although more subjects were employed in this study than is typical in this field, the results require validation.

A number of reports (**4**, **6**, **7**, **8**, **52**, **54**, **58**, **73**, **149**, **159**, **166**, **172**) have related outcome to clinical estimates of personality or behavioral traits. By and large such approaches leave much to be desired in the way of controls and reproducibility. It also appears that among comparable studies there are definite disagreements. Thus Gray (**58**) and Wender (**172**) differ as to whether patients showing maximal participation in group psychotherapy will be most or least likely to improve. Tillotson and Fleming (**52**) indicated that empirical findings may be less reliable than they appear at first glance. After an attempt to confirm the prognostic efficacy of a number of personality traits they had found retrospectively important for chronic alcoholism (**166**), they concluded that "the outcome of treatment of chronic alcoholism has little apparent relation to sociological or personality traits" (**52**, p. **744**).

One of the most "experimental" investigations of this sort is the study of Peters (**121**) in which rating sheets filled out from interviewers' records permitted a frequency count of trait names. Twelve traits were found to be predictive of outcome. Peters interpreted these as showing that a "generally heightened activity level" and a state of "integrity and dominance of the cortical centers" were favorable to improvement. Whether or not this conclusion turns out to be valid, the approach is an admirable attempt to introduce some quantification into what is too often but a potpourri of impressions.

A number of prognostic reports have arisen from studies of the results of nondirective therapy. Snyder (**162**) with only five cases, found the one unsuccessful case expressed fewer feelings and also more negative and ambivalent attitudes toward the counselor than did the four successful cases. Blau (**15**) developed a different attitudinal index, one based on statements referring to the self during the first interview. He found that the greater the number of positively-valenced and ambivalently-valenced self statements, and the fewer the negatively-valenced self statements, the better the prediction for therapeutic success.

There have also been some reports based upon psychoanalytic theory. Knight (**88**) stated that those alcoholic individuals well developed in the "second anal state" offered the best material for therapy. Piers (**122**) claimed that only the "oral types" of schizophrenic responded well to insulin treatment, "anal types" responding poorly. It was tentatively proposed that the decisive psychodynamic factor in insulin treatment was the guiltless fulfillment of an immense oral craving. Obviously, these reports should be considered merely suggestive, since they remain unsupported by evidence or operational definitions.

Skottowe (**160**) has pointed out three types of schizophrenics that he felt called for differently arranged and proportioned forms of therapy. The dys-symbolic type is unable to formulate "conceptual thoughts" upon personal topics or to discriminate linguistically the gradations of his emotions, even though he is in a state of clear consciousness and can use words for perceptual thinking. This type of case, according to Skottowe, fails to recover with shock therapy. The dyskinetic—those with "disorders of motility"—recover well, while the simple paranoid type becomes accessible to further needed psychotherapy through the use of shock therapy. Thomas (**167**) provided data from 32 dys-symbolic cases to support Skottowe's thesis; not one of these cases recovered. A better test of this hypothesis is needed, as well as an enumeration of objective criteria of classification of these three types.

Perhaps the most thorough attempt at determining prognosis has been the work of Wittman with the Elgin scale. This scale contains 30 weighted factors thought to be prognostic for functional psychoses

(**176**). These factors involved numerous psychological and clinical characteristics derived from studies previous to 1941. The scores of schizophrenics were found to be bimodally distributed on this scale, those of manic-depressives skewed (**179**). In the schizophrenic group there was a correlation of prognostic score with diagnostic subcategory: (*a*) the hebephrenic and simple types had poor prognosis; (*b*) the catatonic and undetermined groups were bimodally distributed with regard to prognosis; and (*c*) the paranoid subtype did not fit well into the scale. Prognosis made by the Elgin scale was considerably better than that made by the clinical staff. However, the probability of this advantage occurring by chance was not indicated, the comparison being presented in percentages. Wittman suggested on the basis of her analysis that the duration of psychosis as a criterion for predicting improvement is an artifact rather than a true criterion. Patients with a long duration of illness show poorer prognosis because they are basically ill, not the reverse.

Wittman (**177**) has more recently found a significantly larger number of cases of "constitutional" schizophrenics (those who had been maladjusted as children) than "functional" showed a high positive (unfavorable prognostically) score on the Elgin rating scale, high weighting in heboid regressive features, poor institutional development, and lack of improvement with shock therapy. She thinks that longitudinal studies are necessary for prognosis, and accepts Langfeldt's (**92**) distinction between process schizophrenia and schizophreniform types of psychosis. Data have been reported by Sarbin which further emphasize the usefulness of this scale (**142**). In his study the weights for the scale items were derived empirically, rather than arbitrarily assigned, and similar results were obtained.

There have been quite a few indications of the value of morphological indices for prognosis. Such factors were introduced by Kretschmer's report that dementia praecox patients were most often of asthenic physique and manic-depressives of pyknic physique. This led to the correlative hypothesis that in the functional psychoses pyknics have a better prognosis than do asthenics. Studies that report morphological differences in outcome for mental disorders have usually stressed the better prognosis for patients of pyknic than for those of asthenic habitus (**22, 25, 53, 56, 79, 83, 84, 95, 96, 104, 112, 163**), but some investigators have failed to find any relationship between outcome and body type (**98, 104, 133, 183**) or have found the "pyknotic" unfavorable (**105**).

DISCUSSION

In order for the investigation of prognostic indices to become reasonably effective, studies in this field will have to meet at least four criteria:

1. The conditions of the experiment must be specified. This means that the patient population must be described in all pertinent details, the conditions of therapy must be stipulated, and objective criteria of outcome must be presented. Many of the studies in the literature have failed to specify these crucial characteristics.

2. The experiment must deal with relatively homogeneous populations and conditions. Numerous studies (**106**, **107**, **144**, **145**, **174**, **175**, **181**, **183**) have demonstrated that the prognostic indices applying for groups differing in diagnosis, chronicity, or therapy may differ. The investigation of broad categories such as "psychiatric patients" may be said to be just as meaningful as studies of very delimited groups. But since the differential prognoses of manic-depressives and schizophrenics, for example, are already known, it would appear more useful to discover the prognostic indices for each subgroup. If investigations were carried out under homogeneous conditions we could eventually build up the composite picture, an accomplishment much less likely to result from studies in which these variables are uncontrolled.

3. Findings should be reported in terms amenable to statistical evaluation. Even though most investigators have presented evidence in quantitative form, all too frequently they have not made tests of statistical significance. Such information in terms of probabilities is necessary to establish the confidence that can be put in the findings. Furthermore, tests of significance have often been misinterpreted. Cochran and Cox (**26**) clearly state the misconceptions which can arise from a misapplication of the laws of probability:

> In order that the F- and t-tests be valid, the tests to be made in an experiment should be chosen before the results have been inspected. The reason for this is not hard to see. If tests are selected *after* inspection of the data, there is a natural tendency to select comparisons that appear to give large differences. Now large apparent differences may arise either because there are large real effects, or because of a fortuitous combination of the experimental errors. Consequently, in so far as differences are selected just because they seem to be large, it is likely that an undue proportion of the cases selected will be those where the errors have combined to make the differences large. The extreme case most commonly cited is that of the experimenter who always tests, by an ordinary t-test, the difference between the highest and lowest treatment means. If the number of treatments is large, this difference will be substantial even when the treatments produce no real differences in effect. It may be shown that with 3 treatments the observed value of t will exceed the 5% level in the table about 13% of the time. With 6 treatments the figure is 40%, with 10 treatments 60%, and with 20 treatments 90%. When the experimenter thinks that he is making a t-test at the 5% level, he is actually testing at the 13% level, or the 40% level, and so on (**26**, pp. 67–68).

4. Findings must be subjected to cross validation. Another indication of the lack of statistical sophistication in this field is the large proportion of exploratory findings in relation to the small proportion of confirmatory findings reported in the prognostic literature. Obviously,

exploratory or previously unhypothesized findings must be verified before they can be regarded with confidence.

When these criteria have been met, it will be possible to determine with a much greater degree of accuracy whether psychological tests can have prognostic value and what range of applicability any particular indices may have. Since the majority of prognostic studies have been conducted and interpreted in broad diagnostic, therapeutic, and outcome terms, it has been necessary to evaluate them on this basis. It seems very likely, however, that the broad categories employed in this review will not prove feasible in future prognostic work.

To obtain the prognostically homogenous conditions that seem basic to any future research, there may be a need (possibly common to all fields of psychology) for an over-all coordinating and evaluating organization. Cattell (**24**) has pointed out the need for a quorum vote of policy regarding the factors of mental organization. Others (**50, 76**) have proposed a general psychological research exchange. Similarly, in the field of prognosis a coordinating committee could be of service in organizing more meaningful experiments and encouraging reporting of results in quantitative form.

Thorough discussions of the logical problems involved in prediction and critical considerations of basic research procedures are available (**30, 31, 37, 59, 74, 75, 103, 114, 141, 142**). It would be highly desirable if research workers were more familiar with works of this sort and were guided to a greater extent by the organization they provide.

Summary

There is no dearth of prognostic studies attempting to relate psychological test performance to eventual outcome in psychopathy. However, many of the individual studies fail to demonstrate empirical justification for the conclusions drawn. Further, comparisons among studies reveal little agreement among findings. With the possible exception of certain findings that may prove to be valid, it appears that previous research on prognosis can tell us only that it is not possible to discover reliable prognostic criteria without controlling variables to a much greater extent than has been customary. When investigators in this field recognize the need for certain experimental and evaluative procedures we may hope to discover what, if any, prognostic power psychological tests may have.

BIBLIOGRAPHY

1. Abel, T. M. The Rorschach Test and school success among mental defectives. *Rorschach Res. Exch.*, 1945, **9**, 105–110.
2. Adler, D. L. Normal vs. schizophrenic perception of similarities. *Psychol. Bull.*, 1942, **39**, 507–508. (Abstract)

3. ALBEE, G. W. Patterns of aggression in psychopathology. *J. consult. Psychol.*, 1950, **14**, 465–468.
4. ALBEE, G. W. The prognostic importance of delusions in schizophrenia. *J. abnorm. soc. Psychol.*, 1951, **46**, 208–212.
5. ALBEE, G. W., & GOLDMAN, R. The Picture Frustration Study as a predictor of overt aggression. *J. proj. Tech.*, 1950, **14**, 303–308.
6. ALEXANDER, F., & FRENCH, T. M. *Psychoanalytic therapy.* New York: Ronald, 1946.
7. ARNOT, R., TALBOT, B., & GREENBLATT, M. One to four year follow-up of 205 cases of bilateral prefrontal lobotomy. In M. Greenblatt, R. Arnot, & H. Solomon (Eds.), *Studies in lobotomy.* New York: Grune & Stratton, 1950.
8. BALINSKY, B. Factors in the vocational adjustment of schizophrenics after mental hospital discharge. *J. clin. Psychol.*, 1947, **3**, 341–349.
9. BECK, S. J. Rorschach *F* plus and the ego in treatment. *Amer. J. Orthopsychiat.*, 1948, **18**, 395–401.
10. BELLAK, L. *Dementia praecox.* New York: Grune & Stratton, 1948.
11. BENJAMIN, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. of California Press, 1946. Pp. 65–90.
12. BILLIG, O., & SULLIVAN, D. J. Prognostic data in chronic alcoholism. *Rorschach Res. Exch.*, 1942, **6**, 117–127.
13. BILLIG, O., & SULLIVAN, D. J. Personality structure and prognosis of alcohol addiction. *Quart. J. Stud. Alc.*, 1943, **3**, 554–573.
14. BLAIR, D. Prognosis in schizophrenia. *J. ment. Sci.*, 1940, **86**, 378–477.
15. BLAU, T. H. Report on a method of predicting success in psychotherapy. *J. clin. Psychol.*, 1950, **6**, 403–406.
16. BOLLES, M. M., ROSEN, G. P., & LANDIS, C. Psychological performance tests as prognostic agents for the efficacy of insulin therapy in schizophrenia. *Psychiat. Quart.*, 1938, **12**, 733–737.
17. BOSS, M. Psychopathologie des Traumes bei schizophrenen und organischen Psychosen. *Z. ges. Neurol. Psychiat.*, 1938, **162**, 459–494.
18. BRADWAY, K. P., LION, E. G., & CORRIGAN, H. G. The use of the Rorschach in a psychiatric study of promiscuous girls. *Rorschach Res. Exch.*, 1946, **10**, 105–110.
19. BROWN, M. A. Alcoholic profiles on the Minnesota Multiphasic. *J. clin. Psychol.*, 1950, **6**, 266–269.
20. CARP, A. MMPI performance and insulin shock therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 721–726.
21. CARP, A. Performance on the Wechsler-Bellevue Scale and insulin shock therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 127–136.
22. CARTER, A. B. The prognostic factors of adolescent psychoses. *J. ment. Sci.*, 1942, **88**, 31–81.
23. CATTELL, R. B. Temperament tests in clinical practice. *Brit. J. med. Psychol.*, 1937, **16**, 43–61.
24. CATTELL, R. B. The measurement of adult intelligence. *Psychol. Bull.*, 1943, **40**, 153–193.
25. CHASE, L. S., & SILVERMAN, S. Prognostic criteria in schizophrenia; a critical survey of the literature. *Amer. J. Psychiat.*, 1941, **98**, 360–368.
26. COCHRAN, W. G., & COX, G. M. *Experimental designs.* New York: Wiley, 1950.
27. COHEN, L. H. The therapeutic significance of fear in the metrazol treatment of schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 1349–1357.
28. COOK, L. C. Has fear any therapeutic significance in convulsion therapy? *J. ment. Sci.*, 1940, **86**, 484–490.
29. COWEN, E. L., & COMBS, A. W. Fol-

low-up study of 32 cases treated by nondirective psychotherapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 232–258.

30. Cronbach, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.*, 1949, **46**, 393–429.

31. Dailey, C. A. The clinician and his predictions. *J. clin. Psychol.*, 1951, **7**, 270–273.

32. Davidson, G. M., & Conkey, R. C. The Rorschach Test and the questions of "prognosis" and "recovery" in syphilitic meningo-encephalitis. *Psychiat. Quart.*, 1950, **24**, 243–258.

33. Dearborn, G. V. N. The determination of intellectual regression and progression. *Amer. J. Psychiat.*, 1926-7, **6**, 725–741.

34. Denker, P. G. Results of treatment of psychoneuroses by the general practitioner: a follow-up study of 500 patients. *Arch. Neurol. Psychiat., Chicago*, 1947, **57**, 504–505.

35. Diamond, B. L., & Schmale, H. T. The Mosaic Test. 1. An evaluation of its clinical application. *Amer. J. Orthopsychiat.*, 1944, **14**, 237–250.

36. Dickson, L. P. The Rorschach Test as an indicator of prognosis in psychoneurosis. Unpublished master's thesis, Catholic Univer., 1949.

37. DuMas, F. M. The objective evaluation of prognostic proficiency. *J. gen. Psychol.*, 1950, **43**, 131–135.

38. Earl, C. J. C. A psychograph for morons. *J. abnorm. soc. Psychol.*, 1940, **35**, 428–448.

39. Ellis, R. W., & Brown, G. G. The nature of Rorschach responses from pulmonary tuberculosis patients. *J. clin. Psychol.*, 1950, **6**, 298–300.

40. Farberow, N. L. Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, **42**, 3–79.

41. Feldman, M. J. A prognosis scale for shock therapy. *Amer. Psychologist*, 1948, **3**, 348. (Abstract)

42. Feldman, M. J. Some scales to predict and evaluate the results of several types of shock therapy. Unpublished doctoral dissertation, Univer. of California, 1949.

43. Fiedler, F. E., & Siegel, S. M. The Free Drawing Test as a predictor of non-improvement in psychotherapy. *J. clin. Psychol.*, 1949, **5**, 386–389.

44. Filmer-Bennett, G. T. A study of prognostic indices in the Rorschach records of hospitalized patients. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1950.

45. Fisher, K. A. Changes in test performance of ambulatory depressed patients undergoing electro-shock therapy. *J. gen. Psychol.*, 1949, **41**, 195–232.

46. Fisher, S. Rorschach patterns in conversion hysteria. *J. proj. Tech.*, 1951, **15**, 98–108.

47. Fisher, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, **15**, 250–254.

48. Fisher, S., & Fisher, Rhoda. Value of isolation rigidity in maintaining integration in seriously disturbed personalities. *J. Pers.*, 1950, **19**, 41–47.

49. Fishman, S. An investigation of the relationship between references to the self-concept of above the knee amputees, and their adjustment to leg prostheses. *Amer. Psychologist*, 1949, **4**, 256. (Abstract)

50. Fiske, D. W. Must psychologists be experimental isolationists? *Amer. Psychologist*, 1947, **2**, 23, 28.

51. Flanagan, J. C. (Ed.) The Aviation Psychology Program in the Army Air Forces. *AAF Aviat. Psychol. Res. Rep.* No. 1. Washington: U. S. Government Printing Office, 1948.

52. Fleming, R., & Tillotson, K. Further studies on the personality and sociological factors in the prognosis and treatment of chronic alcohol-

ism. *New Engl. J. Med.*, 1939, **221**, 741–745.

53. GILDEA, E. F., & MAN, E. B. Methods of estimating capacity for recovery in patients with manic-depressive and schizophrenic psychoses. *Amer. J. Psychiat.*, 1943, **99**, 496–506.

54. GLATZER, H. T. Selection of mothers for group therapy. *Amer. J. Orthopsychiat.*, 1947, **17**, 477–483.

55. GOLDSTEIN, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **96**, 575–588.

56. GOTTLIEB, B. S. Prognostic criteria in hebephrenia; the importance of age, sex, constitution and marital status. *Amer. J. Psychiat.*, 1940, **97**, 332–341.

57. GRAHAM, V. T. Psychological studies of hypoglycemia therapy. *J. Psychol.*, 1940, **10**, 327–358.

58. GRAY, W. Group psychotherapy in a state hospital. *J. nerv. ment. Dis.*, 1948, **108**, 485–495.

59. GUILFORD, J. P., & MICHAEL, W. B. *The prediction of categories from measurements: with application of personnel selection and clinical prognosis.* Ann Arbor: Edwards Bros., 1949.

60. HALES, W. M., & SIMON, W. Minnesota Multiphasic Personality Inventory patterns before and after insulin shock therapy. *Amer. J. Psychiat.*, 1948, **105**, 254–258.

61. HALPERN, F. Rorschach interpretation of the personality structure of schizophrenics who benefit from insulin therapy. *Psychiat. Quart.*, 1940, **14**, 826–833.

62. HAMLIN, R. Test pattern of high grade mentally defective girls. *Proc. Amer. Ass. ment. Def.*, 1938, **43**, 161–165.

63. HARRIS, I. D., RAPOPORT, L., & RYNERSON, M. A. Observations on asthmatic children. *Amer. J. Orthopsychiat.*, 1950, **20**, 490–505.

64. HARRIS, R. E. Measured personality characteristics of convulsive therapy patients: a study of diagnostic and prognostic criteria. *Psychol. Bull.*, 1945, **42**, 535. (Abstract)

65. HARRIS, R. E. Psychodiagnostic testing in psychiatry and psychosomatic medicine. In M. R. Harrower (Ed.), *Recent advances in diagnostic psychological testing.* Springfield: Thomas, 1950. Pp. 13–29.

66. HARRIS, R. E., BOWMAN, K. M., & SIMON, A. Studies in electronarcosis therapy. III. Psychological test findings. *J. nerv. ment. Dis.*, 1948, **107**, 371–376.

67. HARRIS, R. E., & CHRISTIANSEN, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, **21**, 269–284.

68. HARTMAN, A. A. An experimental examination of the Thematic Apperception Technique in clinical diagnosis. *Psychol. Monogr.*, 1949, **63**, No. 8 (Whole No. 303).

69. HERTZ, M. R. The role of the Rorschach method in planning for treatment. *Rorschach Res. Exch.*, 1945, **9**, 134–146.

70. HERTZ, M. R. The Rorschach method and its significance in the Mental Hygiene Program. In P. L. Harriman (Ed.), *Twentieth century psychology.* New York: Philosophical Library, 1946. Pp. 652–684.

71. HERTZ, M. R. Suicidal configurations in Rorschach records. *Rorschach Res. Exch.*, 1948, **12**, 3–58.

72. HERTZ, M. R. Further study of "suicidal" configurations in Rorschach records. *Rorschach Res. Exch.*, 1949, **13**, 44–73.

73. HOCH, P. H. Personality factors in alcoholic psychosis. *Psychiat. Quart.*, 1940, **14**, 338–346.

74. HORST, P. *The prediction of personal adjustment.* New York: Soc. Sci. Res. Coun., Bull. 48, 1941.

75. HOTELLING, H., SOROTKIN, P., GUTT-

MAN, L., & BURGESS, E. W. The prediction of personal adjustment: A symposium. *Amer. J. Sociol.*, 1942, **48**, 61–86.

76. IRWIN, F. W. A proposal toward a psychological research exchange. *Amer. Psychologist*, 1948, **3**, 91–94.

77. KALINOWSKY, L. B., & HOCH, P. H. *Shock treatments and other somatic procedures in psychiatry.* New York: Grune & Stratton, 1946.

78. KAMMAN, G. R. The Rorschach method as a therapeutic agent. *Amer. J. Orthopsychiat.*, 1944, **14**, 21–27.

79. KANT, O. The evaluation of prognostic criteria in schizophrenia. *J. nerv. ment. Dis.*, 1944, **100**, 598–605.

80. KATZ, H. Untersuchungen an insulinbehandelten Schizophrenen mit dem Rorschach'schen Formdeutversuch. *Mschr. Psychiat. Neurol.*, 1941, **104**, 15–33.

81. KENDIG, ISABELLE V. Projective techniques as a psychological tool in diagnosis. *J. clin. Psychopath.*, 1944, **6**, 101–110.

82. KISKER, G. W. A projective approach to personality patterns during insulin-shock and metrazol-convulsive therapy. *J. abnorm. soc. Psychol.*, 1942, **37**, 120–124.

83. KLINE, N. S., & TENNEY, A. M. Constitutional factors in the prognosis of schizophrenia. *Amer. J. Psychiat.*, 1950, **107**, 434–441.

84. KLINE, N. S., & TENNEY, A. M. Prognosis in topectomies and lobotomies relative to body type. *Arch. Neurol. Psychiat., Chicago*, 1951, **65**, 323–325.

85. KLOPFER, W. G. The efficacy of group therapy as indicated by group Rorschach records. *Rorschach Res. Exch.*, 1945, **9**, 207–209.

86. KLUGMAN, S. F. Test scores of psychoneurotic returnees before and after convalescent furlough. *J. clin. Psychol.*, 1948, **4**, 195–200.

87. KLÜVER, H. Psychology at the beginning of World War II: Meditations on the impending dismembership of psychology written in 1942. *J. Psychol.*, 1949, **28**, 383–410.

88. KNIGHT, R. P. The dynamics and treatment of chronic alcohol addiction. *Bull. Menninger Clin.*, 1937, **1**, 233–250.

89. KRIEGMAN, G., & HILGARD, J. R. Intelligence level and psychotherapy with problem children. *Amer. J. Orthopsychiat.*, 1944, **14**, 251–265.

90. KRUGMAN, M. Rorschach examination in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1941, **11**, 503–511.

91. LANDIS, C., & ERLICK, D. An analysis of the Porteus Maze-Test as affected by psychosurgery. *Amer. J. Psychol.*, 1950, **63**, 557–566.

92. LANGFELDT, G. *The prognosis in schizophrenia and the factors influencing the course of the disease.* London: Oxford Univer. Press, 1937.

93. LEVI, J. Rorschach patterns as a tool in predicting success or failure in the rehabilitation of the physically handicapped. *Amer. Psychologist*, 1950, **5**, 320–321. (Abstract)

94. LEVI, J. Rorschach patterns predicting success or failure in the rehabilitation of the physically handicapped. *J. abnorm. soc. Psychol.*, 1951, **46**, 240–244.

95. LEWIS, A. J. Prognosis in schizophrenia. *Lancet*, 1935, **228**, 339–341.

96. LEWIS, N. D. C. What do we know about dementia praecox? *Ment. Hyg.*, 1950, **34**, 569–581.

97. LORR, M., WITTMAN, PHYLLIS, & SCHANBERGER, W. An analysis of the Elgin Prognostic Scale. *J. clin. Psychol.*, 1951, **7**, 260–263.

98. LOURIE, R. S., PACELLA, B. L., & PIOTROWSKI, Z. A. Studies on the prognosis in schizophrenic-like psychoses in children. *Amer. J. Psychiat.*, 1943, **99**, 542–551.

99. Low, A. A., Singer, H. D., & McCorry, C. L. Life situation tests as aids in psychiatric prognosis. *Amer. J. Psychiat.*, 1939, **96**, 147–162.

100. Low, A. A., Sonenthal, I. R., Blaurock, M. F., Kaplan, M., & Sherman, I. Metrazol shock treatment of the "functional" psychoses. *Arch. Neurol. Psychiat., Chicago*, 1938, **39**, 717–736.

101. Luborsky, L. B. Psychometric changes during electric shock treatment. *J. nerv. ment. Dis.*, 1948, **107**, 531-536.

102. McCall, R. J. Psychometric evaluation of Rorschach records in brain-operated patients. Unpublished doctoral dissertation, Columbia Univer., 1951.

103. Malamud, D. I. Objective measurement of clinical status in psychopathological research. *Psychol. Bull.*, 1946, **43**, 240–258.

104. Malamud, W., & Gottlieb, J. S. Course and prognosis in the psychoneuroses. *J. Mt. Sinai Hosp., New York*, 1942, **9**, 630–639.

105. Malamud, W., & Render, N. Course and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 1039–1057.

106. Malamud, W., Sands, S. L., & Malamud, I. The involutional psychoses: a socio-psychiatric study. *Psychosom. Med.*, 1941, **3**, 410–426.

107. Malamud, W., Sands, S. L., Malamud, I. T., & Powers, P. J. P. The involutional psychoses: a socio-psychiatric follow-up study. *Amer. J. Psychiat.*, 1949, **105**, 567–572.

108. Masserman, J. H., & Balken, E. R. The clinical application of phantasy studies. *J. Psychol.*, 1938, **6**, 81–88.

109. Menninger, W. C. The treatment of chronic alcohol addiction. *Bull. Menninger Clin.*, 1938, **2**, 101–112.

110. Miles, H. H. W., Barrabee, E. L., & Finesinger, J. E. Evaluation of psychotherapy, with a follow-up study of 62 cases of anxiety neurosis. *Psychosom. Med.*, 1951, **13**, 83–105.

111. Morris, W. W. Prognostic possibilities of the Rorschach method in metrazol therapy. *Amer. J. Psychiat.*, 1943, **100**, 222–230.

112. Mueller, M. Prognosis, schizophrenic diseases. *Fortschr. Neurol. Psychiat.*, 1935, **7**, 447–474.

113. Muench, G. A. The application of diagnostic psychological methods to counseling and psychotherapy. In M. R. Harrower (Ed.), *Recent advances in diagnostic psychological testing*. Springfield, Ill.: Thomas, 1950. Pp. 53–72.

114. Myers, C. R. Prediction in clinical psychology. *Canadian J. Psychol.*, 1950, **4**, 97–108.

115. Nagle, J. M. Two factors in the prognosis of alcoholism. *Psychiat. Quart.*, 1942, **16**, 633–640.

116. O'Connell, J. J., & Penrose, L. S. Tests of psychomotor efficiency in patients treated with metrazol. *J. ment. Sci.*, 1941, **87**, 183–191.

117. Oppenheim, S., & Brower, D. The effects of electric shock therapy as revealed by the Rorschach technique. *Psychiat. Quart.*, 1948, **22** (supplement), 318–325.

118. Pacella, B. L., Piotrowski, Z., & Lewis, N. D. C. The effects of electric convulsive therapy on certain personality traits in psychiatric patients. *Amer. J. Psychiat.*, 1947, **104**, 83–91.

119. Pearson, J. S. Prediction of the response of schizophrenic patients to electro-convulsive therapy. *J. clin. Psychol.*, 1950, **6**, 285–287.

120. Pearson, J. S., & Swenson, W. M. A note on extended findings with the MMPI in predicting response to electro-convulsive therapy. *J. clin. Psychol.*, 1951, **7**, 288.

121. Peters, H. N. Traits related to improved adjustment of psychotics after lobotomy. *J. abnorm. soc. Psychol.*, 1947, **42**, 383–392.

122. Piers, G. Prognostic observations in insulin treatment of schizophrenia. *Elgin Pap.*, 1941, **4**, 34–45.

123. Piotrowski, Z. The prognostic possibilities of the Rorschach method in insulin treatment. *Psychiat. Quart.*, 1938, **12**, 679–689.

124. Piotrowski, Z. Rorschach manifestations of improvement in insulin treated schizophrenics. *Psychosom. Med.*, 1939, **1**, 508–526.

125. Piotrowski, Z. A simple experimental device for the prediction of outcome of insulin treatment in schizophrenia. *Psychiat. Quart.*, 1940, **14**, 267–273.

126. Piotrowski, Z. The Rorschach method as a prognostic aid in the insulin shock treatment of schizophrenics. *Psychiat. Quart.*, 1941, **15**, 807–822.

127. Piotrowski, Z. The modifiability of personality as revealed by the Rorschach method: methodological considerations. *Rorschach Res. Exch.*, 1942, **6**, 160–167.

128. Piotrowski, Z., & Lewis, N. D. C. An experimental criterion for the prognostication of the status of schizophrenics after a three-year interval based on Rorschach data. In P. H. Hoch and J. Zubin (Eds.), *Relation of psychological tests to psychiatry.* New York: Grune & Stratton, 1951. Pp. 51–72.

129. Rabin, A. I. Homicide and attempted suicide: A Rorschach study. *Amer. J. Orthopsychiat.*, 1946, **16**, 516–524.

130. Rabin, A. I. Effects of electric shock treatment upon some aspects of personality and intellect. *Amer. Psychologist*, 1947, **2**, 284. (Abstract)

131. Rees, W. L., & Jones, A. M. An evaluation of the Rorschach Test as a prognostic aid in the treatment of schizophrenia by insulin coma therapy, electronarcosis, electroconvulsive therapy and leucotomy. *J. ment. Sci.*, 1951, **97**, 681–689.

132. Reiman, M. G. The prognostic value of mental symptoms in the psychoses. *Stud. Psychol. Psychiat.*, 1935, **3**, No. 6. Pp. 40.

133. Rennie, T. A. C. Analysis of one hundred cases of schizophrenia with recovery. *Arch. Neurol. Psychiat., Chicago*, 1941, **46**, 197–229.

134. Rhine, J. B. ESP and prognosis. *J. Insur. Med.*, 1949, **4**, 16–17.

135. Rioch, M. J. The use of the Rorschach Test in the assessment of change in patients under psychosurgery. *Psychiatry*, 1949, **12**, 427–434.

136. Roberts, L. K. The predictive significance of the Rorschach Technique in relation to the outcome of psychotherapy. Unpublished doctoral dissertation, Univer. of Iowa, 1950.

137. Ruesch, J., *et al.* Chronic disease and psychological invalidism. *Psychosom. Med. Mon.*, No. 9, New York, 1946.

138. Ruesch, J., & Bowman, K. M. Prolonged post-traumatic syndromes following head injury. *Amer. J. Psychiat.*, 1945, **102**, 145–163.

139. Ruesch, J., Christiansen, C., Patterson, L. C., Dewees, S., & Jacobson, A. Psychological invalidism in thyroidectomized patients. *Psychosom. Med.*, 1947, **9**, 77–91.

140. Salfield, D. J. The usefulness of the Rorschach Test for diagnosis, prognosis and epicrisis, mainly in child guidance treatment. *J. ment. Sci.*, 1951, **97**, 84–89.

141. Sarason, S. S. The test situation and the problem of prediction. *J. clin. Psychol.*, 1950, **6**, 387–392.

142. Sarbin, T. R. The logic of prediction in psychology. *Psychol. Rev.*, 1944, **51**, 210–228.

143. Scherer, I. W. Correlates of improvement with electro convulsive therapy. *Psychol. Serv. Cent. J.*, 1949, **1**, 109–121.

144. Scherer, I. W. Prognoses and psychological scores in electroconvulsive therapy, psychosurgery, and spontaneous remission. *Amer. J. Psychiat.*, 1951, **107**, 926–931.

145. Schnack, G. F., Shakow, D., & Lively, M. L. Studies in insulin and metrazol therapy: I. The differential prognostic value of some psychological tests. *J. Pers.*, 1946, **14**, 106–124.

146. Schnack, G. F., Shakow, D., & Lively, M. L. Studies in insulin and metrazol therapy: II. Differential effects on some psychological functions. *J. Pers.*, 1946, **14**, 125–149.

147. Schofield, W. Research in clinical psychology: 1949. *J. clin. Psychol.*, 1950, **6**, 234–237.

148. Schofield, W. Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. *Psychol. Monogr.*, 1950, **64**, No. 5 (Whole No. 311).

149. Seitz, P. F. D. A dynamic factor correlated with the prognosis in paranoid schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1951, **65**, 604–606.

150. Sexton, M. C. The Autokinetic Test: Its value in psychiatric diagnosis and prognosis. *Amer. J. Psychiat.*, 1945, **102**, 399–402.

151. Shaskin, D. A., Conrad, D. C., & Grant, J. D. Prediction of behavior in group psychotherapy from Rorschach protocols (preliminary report). *Group Psychotherapy*, 1950, **3**, 218–230.

152. Siegel, M. G. The Rorschach Test as an aid in selecting clients for group therapy and evaluating progress. *Ment. Hyg.*, 1944, **28**, 444–449.

153. Siegel, M. G. The use of the Rorschach Test in a treatment program. *Rorschach Res. Exch.*, 1945, **9**, 126–129.

154. Siegel, M. G. The Rorschach Test in diagnosis and prognosis. *Family*, 1946, **27**, 102–106.

155. Siegel, M. G. The diagnostic and prognostic validity of the Rorschach Test in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, **18**, 119–133.

156. Siegel, S. M. Personality factors in psychotherapeutic improvement as identified by the Rorschach Test. *Amer. Psychologist*, 1951, **6**, 341–342. (Abstract)

157. Sillman, L. R. Chronic alcoholism. *J. nerv. ment. Dis.*, 1948, **107**, 127–149.

158. Simon, W., & Hales, W. Note on a suicide key in the Minnesota Multiphasic Personality Inventory. *Amer. J. Psychiat.*, 1949, **106**, 222–223.

159. Simon, B., & Holt, W. L., Jr. The relief of specific psychiatric symptoms by convulsive shock therapy. *Dis. nerv. Syst.*, 1946, **7**, 241–244.

160. Skottowe, I. Shock therapy; a plea for proportion in psychiatry. *Proc. roy. Soc. Med.*, 1939, **32**, 843–852.

161. Sloan, W. Prediction of extramural adjustment of mental defectives by use of the Rorschach test. *J. consult. Psychol.*, 1948, **12**, 303–309.

162. Snyder, W. U. A comparison of one unsuccessful with four successful nondirectively counseled cases. *J. consult. Psychol.*, 1947, **11**, 38–42.

163. Ssucharewa, G. E., & Kogan, E. I. Prognosis of schizophrenia of childhood and puberty. *Sovetsk. psikhonevr.*, 1933, **6**, 120–131.

164. Stainbrook, E. Shock therapy: Psychologic theory and research. *Psychol. Bull.*, 1946, **43**, 21–60.

165. Terry, G. C., & Rennie, T. A. C. Analysis of parergasia. *Nerv. ment. Dis. Monogr. Ser.*, No. 64, 1938.

166. Tillotson, K., & Fleming, R. Personality and sociologic factors in the prognosis and treatment of chronic alcoholism. *New Engl. J. Med.*, 1937, **217**, 611–615.

167. THOMAS, J. S. Some clinical examples of "dys-symbole": its relation to shock therapy. *J. ment. Sci.*, 1940, **86**, 100–108.

168. VAN BARK, B., & BARON, S. Neurotic elements in the Rorschach records of psychotics. *Rorschach Res. Exch.*, 1943, **7**, 166–168.

169. VOTH, A. C. An experimental study of mental patients through the autokinetic phenomenon. *Amer. J. Psychiat.*, 1947, **103**, 793–805.

170. WALL, C. Observations on the behavior of schizophrenic patients undergoing insulin shock therapy. *J. nerv. ment. Dis.*, 1940, **91**, 1–8.

171. WECHSLER, D., HALPERN, F., & JAROS, E. Psychometric study of insulin-treated schizophrenics. *Psychiat. Quart.*, 1940, **14**, 466–476.

172. WENDER, H. Experiences in group psychotherapy with insulin-treated patients. *Psychiat. Quart.*, 1950, **24**, 314–323.

173. WHEELER, E. O., WHITE, P. D., REED, E. W., & COHEN, M. E. Neurocirculatory asthenia (anxiety, neurosis, effort syndrome, neurasthenia). *J. Amer. med. Ass.*, 1950, **142**, 878–889.

174. WILLENSON, D. Relation of subtest performance on the Wechsler-Bellevue Intelligence Scale to the effects of psychosurgery. Unpublished master's thesis, Columbia Univer., 1950.

175. WINDLE, C., & HAMWI, V. An exploratory study of the prognostic value of the Complex Reaction Time Test in acute and chronic psychotics. To be published.

176. WITTMAN, PHYLLIS. A scale for measuring prognosis in schizophrenic patients. *Elgin Pap.*, 1941, **4**, 20–33.

177. WITTMAN, PHYLLIS. Diagnostic and prognostic significance of the shut-in personality type as a prodromal factor in schizophrenia. *J. clin. Psychol.*, 1948, **4**, 211–214.

178. WITTMAN, PHYLLIS, & RUSSELL, J. T. Mental efficiency levels before and after shock therapy. *J. gen. Psychol.*, 1942, **26**, 3–17.

179. WITTMAN, PHYLLIS, & STEINBERG, L. Follow-up of an objective evaluation of prognosis in dementia praecox and manic-depressive psychoses. *Elgin Pap.*, 1944, **5**, 216–227.

180. ZUBIN, J. Discussion I. In P. H. Hoch and J. Zubin (Eds.), *Relation of psychological tests to psychiatry.* New York: Grune & Stratton, 1951. Pp. 96–104.

181. ZUBIN, J., & THOMPSON, JANE. *Sorting tests in relation to drug therapy in schizophrenia.* New York: New York State Psychiatric Institute, 1941.

182. ZUBIN, J., & WINDLE, C. The prognostic value of the Metenym Test in a follow-up study of psychosurgery patients and their controls. *J. clin. Psychol.*, 1951, **7**, 221–223.

183. ZUBIN, J., WINDLE, C., & HAMWI, V. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. To be published.

Received February 18, 1952.

SOME OBSERVATIONS ON *Q* TECHNIQUE

WILLIAM STEPHENSON
University of Chicago

With the exception of the present author, all who have given some consideration to *R* and *Q*[1] techniques have decided that, after all, they are not different in any important respect. Thomson (**32**) seems undecided, but Burt (**3**), Cattell (**4**), Loevinger (**19**), Babington-Smith (**15**) and others argue that the same data can be examined both ways, by *R* or *Q*, with the same results. If there are more persons than tests, then *tests* are correlated (*R*), but if there are more tests than persons, then *persons* are correlated instead (*Q*). All that distinguishes the two at bottom, therefore, seems to be a matter of convenience. Our own view, first given in 1935 (**25**), is that such arguments are purely superficial, and in no way represent what is at issue. A little logical analysis indicated at once, so it seemed to us, that wholly distinct principles were involved in *R* and *Q* respectively, leading to quite different *systems* (**27**) as we first named them, that is to different *methodologies* as they are now called. Burt and the present writer, in particular, agreed to differ about these issues (**3**).

However, the position is now much clearer than it was in 1935, thanks to advances in modern logical analysis and to the influence of Fisher's (**9**) methodology. It is now certain that not only are *R* and *Q* as distinctive as we said they were, but also that the former is based upon many fallacies which are obviated in *Q*. It is the purpose of the present paper to outline some of these matters.

What is at issue is roughly as follows. When Spearman introduced factor analysis it appeared to offer unusual promise. But serious mistakes were made at its birth. It was linked to interdependency analysis (**15**) and, consequently, to statistical speculations about "factors of the mind," "unitary factors," "primaries," and the like. Sight was lost altogether of quite a different possibility, that factor analysis might serve, instead, merely as an adjunct to dependency analysis, in which the concern is with psychological experiments and not with statistical interdependencies. We defined *Q*, originally, merely as an experimental adjunct (**26**, **27**). Similarly, because of wrong conceptions about scientific method, it was protopostulatory to *R* that psychological theories are the same thing, methodologically, as general propositions, which are testable for their "general implications." We now know that this is not only unsound in principle, but that it could do no other than stultify all attempts to explore psychological theories in their own right, with respect to the correct logic of singular propositions derived from the theories (**14**, **23**). It is essentially this latter logic of scientific method that Q achieves, and its promise for psychology, consequently, is not less than we claimed for it originally. In reintroducing Q methodology, therefore, we have the double task of adjusting many far-reaching mistakes of the early factorists, and of giving a foretaste of the wide applications of factor analysis as an adjunct to experimental method.

---

[1] Cattell (**4**) has added other letters of the alphabet to these two, but his various applications stem, methodologically, from the four *systems* originally defined by Stephenson (**27**), and, as the present paper will indicate, the additional letters have reference to *variate-designs* and not to *methodologies* of the order we are to discuss.

EARLY FORMULATIONS

The applications of factor analysis that are best known, of course, are those with which Spearman began his work, now called R technique, in which the individual differences provided by *tests* or the like are correlated and factored. R technique is a technical way of studying individual differences for attributes of persons. Q technique was defined in 1935 (**27**) as a *system* or *methodology*, radically different from that for R. Clearly others, including Beebe-Center (**1**) and Burt (**3**), had correlated persons long before Thomson (**32**) and Stephenson (**26**) drew attention to these systematic possibilities, but the grasp of this matter in its widest methodological respects was overlooked prior to our early papers (**25**, **26**), and has been missed up to now by many who have sought to expound Q technique, including Burt (**3**), Cattell (**4**), Babington-Smith (**15**), and Loevinger (**19**).

We defined R and Q as two independent systems, the one concerned with individual differences by postulation, and the other not. Burt (**3**) offered a proof that the two were related, but his *reciprocity principle* merely dealt with any one matrix, doubly centered, and with covariance analysis, and therefore touched upon neither R nor Q, much less any conceivable relation between them (**32**). By postulation, R and Q always involve *two* quite different, and singly centered, tables of correlations, each subserved by its own distinctive quantitative and qualitative principles. It is therefore a mistake to argue as though all that is involved is a single matrix of data which, when correlated down the rows is R, and along the columns is Q. Quantitative principles for Q were described in 1935, and were perhaps never taken seriously enough by our critics. They may be summarized briefly as follows: (*a*) The populations are statements, traits, or the like; (*b*) variates refer to operations of a single person, or about him, in one interactional setting; (*c*) the transitory postulate (**16**) has reference to intra-individual differences of "significance"; (*d*) variates may interact; (*e*) scores are approximately normal and standardized, as in product-moment correlational theory generally; (*f*) all the important information for each array is contained in its variation (no information is contained in the variate means); (*g*) the operations of, or about, a person are all subject to the principle of randomization (**9**); and (*h*) the concern is with dependency analysis.

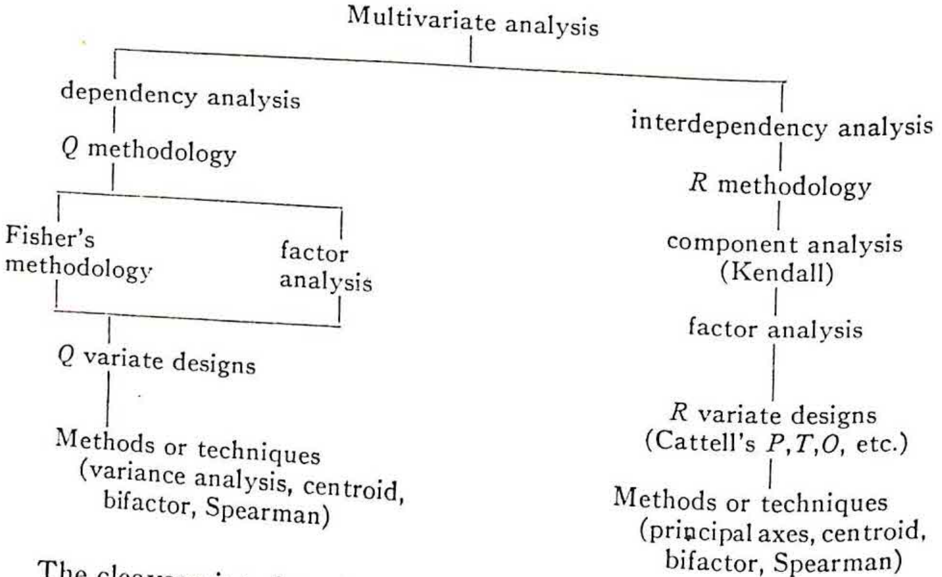
The quantitative principles for *R* concern persons as populations, attributes as variates, and the transitory postulate is made to work in terms of individual differences; the variates (tests, etc.) in *R* do not interact—operations are all subject to the "rule of the single variable"; nor do persons interact; and the concern is with *interdependency*, and not with *dependency* forms of analysis. The qualitative principles are also very different for *R* and *Q* techniques. Those for *R* technique have been based upon an essentially inductive methodology, and upon the study of psychological theories for their "general implications." In *Q* methodology, on the contrary, the concern is with a postulatory-dependency methodology, and with psychological theories as growing-points for singular testable propositions. We believe that the former methodology has had seriously restrictive effects upon psychology, whereas the latter is in keeping with the modern logic of scientific method, and offers almost unlimited scope for an experimental approach to psychology.

We must introduce, therefore, a number of basic formulations which have hitherto had little attention in journals. They are simple enough matters, however, such as (*a*) employing an alternative to the classical, large-sample doctrine of sampling a parent population, namely one which draws upon small-sample doctrines and Fisher's methodology (**9**), (*b*) replacing interdependency analysis by a postulatory-dependency methodology, and (*c*) drawing no arbitrary distinctions between what is "objective" and what is "subjective" in psychological science.

Q AND *R* AS METHODOLOGIES

Q technique is not another method of factor analysis in the sense that the centroid, bifactor, and principal axes methods are distinguishable. These and other methods or techniques, including variance design and analysis, subserve both *Q* and *R* methodologies. Thus it has been a source of some confusion to think of either *R* or *Q* as a mere method or technique in any narrow sense. Each is a complex methodology, involving its own principles, and within which any of the current methods or techniques of statistical analysis, including factor, variance, probability analysis, and the like, may be applied. But, up to now factor analysis has been regarded widely as a matter of interdependency analysis (**3, 15**); almost everyone believes that the concern is with the discovery of factors as unitaries (**32**) or primaries (**33**). An inductive methodology is at issue in which factors have first to be found, and afterwards interperted (**33**). Our own standpoint in *Q* methodology is very different from this: we postulate hypotheses, explanations, and interpretations at the outset in relation to a psychological theory; a set of propositions is then asserted which is put to an empirical test. To the latter end (*a*) structured *Q* samples are composed, where possible, which entail the independencies of the theory at issue, or, if unstructured samples are used, the independencies are implied, and (*b*) variate-designs are employed, with the object of bringing dependencies to light, that is, such as put the propositions to test. *Q* technique, in a narrow sense, consists of experimenting with these two kinds of designs (*a*) and (*b*); factors represent dependent effects and, in certain circumstances, the possibility of lawfulness and underlying tendencies. The effects are tested for significance by any statistical methods that may be appropriate to the situation, variance analysis (Fisher) and centroid factor procedures (Thurstone) being the most immediately useful. Thus, *Q* methodology is concerned with testing concrete, singular, propositional sets.

Following Kendall's example (**15**) it may be of some assistance to offer a kind of genealogical tree for the situation vis-a-vis *R* and *Q* as methodologies:


Multivariate analysis
dependency analysis
Q methodology
Fisher's methodology
factor analysis
Q variate designs
Methods or techniques (variance analysis, centroid, bifactor, Spearman)
interdependency analysis
R methodology
component analysis (Kendall)
factor analysis
R variate designs (Cattell's P,T,O, etc.)
Methods or techniques (principal axes, centroid, bifactor, Spearman)

The cleavage into interdependency and dependency forms of analysis is indicated. In the former case no variate is picked out for prior or any special regard (**15, 34**), whereas in the latter the concern is with independent and dependent variables. The variate designs can be legion, far exceeding any number of letters of the alphabet; the various so-called systems of Cattell (4) have the status of such variate designs. The joining of Fisher's methodology to factor analysis is indicated for *Q* methodology. What is involved in dependency analysis has always been regarded as the essence of experimental method, namely, that independent variables or independencies should be specified, and, as a result of operations in relation to them, dependent effects should be obtained. Usually only one dependent effect is studied at a time, but in Fisher's methodology many may be so studied contemporaneously. We seek to achieve such a methodology, then, for *Q*. The greatest single aid to this end is in the use of structured *Q* samples, a matter to which we shall now attend.

ON STRUCTURED SAMPLES

Large sampling conditions are at issue in *R* and small sampling conditions in *Q*. First a distinction is drawn between populations and statistical universes (**20**). A sample of 200 children in *R* is a population sample, but the scores of these children on a test constitute a statistical distribution. For any one person-population there can be innumerable statistical universes. When large sampling conditions are strictly adhered to, area, stratified, biased, controlled and similar devices are employed in order to reach representative sampling or to allow for departures from it. Our particular innovation has reference to population samples: we propose that these samples can be structured with respect to specifiable independencies, for balanced block or other Fisherien designs (**9**). The general formulation is as follows:

If there can be defined certain independencies *A*, *B*, *C*, with "levels" *a*, *b*, *c* . . . respectively, then, without replication, but with a balanced block design, there are $a \times b \times c$ combinations, one "level" at a time for each independency. If these are replicated *m* times, there will be $m \times abc$ such combinations. We

TABLE 1

FACTORIAL DESIGN FOR A SAMPLE OF PERSONS

Independencies	*Levels*			*No.*	*Degrees of freedom*
W. Age	(*a*) 15–20 years		(*b*) 20–30 years	2	1
X. Socioeconomic status	(*c*) A	(*d*) B	(*e*) CDE	3	2
Y. Educational status	(*f*) University		(*g*) High School	2	1
Z. Habitat	(*h*) Rural		(*i*) City	2	1

aesthetics, and for personality theory in general. Thus, in order to represent Rorschach's theory for his projective technique, the design in Table 3 is appropriate. The design leads to $3 \times 3 \times 2 = 18$ combinations without replication, and may be covered by Rorschach indicators of the kind $(afh) = FC > CF + C$, or by statements of the inferential kind that have reference to such indicators, e.g., "sublimating" (*bfg*). We can also represent in this way conclusions that have been drawn from previous studies with respect to presumed independencies.

THE POSTULATORY-DEPENDENCY METHODOLOGY

The structured samples arrived at in the above manner are not testable propositions in terms of which we seek to prove or disprove a theory or hypothesis by validating the postulated independencies for their "general implications." That is, there is no question of *proving* that the statements in *Q*, any more than the persons in *R*, are as they are asserted to be "on the average," or "in general," or as "indicated from the study of individual differences." The *Q* samples are used, instead, for singular tests of propositions that the theory suggests, or that are deduced from it.

The methodology is not restricted to broad theories, but it is easiest to discuss it in terms of theories such as Spearman propounded about noegenesis (**24**), Jung about personality, Freud about the unconscious, or Rogers about the self. These would probably be regarded as unscientific per se or *in esse* from the *R* methodology standpoint; or, if a so-called scientific approach were made to them along *R* technique lines, as was attempted for the theories of Jung and Spearman, it would be assumed that general propositions are at issue, to be examined for their "general implications." This confuses a rational theory, however, with a general or universal proposition. Theories should be regarded, instead, as growing-points for singular propositions (**14**, **23**). Thus, many attempts have been made in the past to measure introversion-extroversion, based on the supposition that all persons are introverts or extroverts in some degree habitually. The proposition is characteristic of all *R* technique studies, since everyone is supposed to have all attributes in some degree. The Guilfords (**10**) and others, as is well known, found several factors instead of one, such as (they supposed) Jung's theory adumbrated: but each *R* factor has the same postulation at its roots—all persons have each in some degree. This indeed seems inescapable and self-evident.

However, such general implications are nowhere necessarily involved in the theory. The concern, instead, can be with singular propositions, of the following kind.

P.1: *This particular person X is either introvert or extrovert habitually, or neither.*

This we can at once test along Q lines, in terms of the *theory* of introversion-extroverison, without reference to any person other than X, and without the use of any norms or standard scales, and without operational reference to any individual differences. We merely invite X to offer a self-description of himself, as he conceives himself to be habitually, in terms of a Q sort, for a structured sample based on Table 2. In a particular case, for example, we had available a sample of 160 statements, i.e., for 10 replications of the design of Table 2. Our landlady was invited to give a self-description,[2] for the following frequency distribution of scores:

	Most characteristic of me habitually						*Least characteristic of me habitually*		
Score	8	7	6	5	4	3	2	1	0
Frequency	8	12	20	24	32	24	20	12	8

$n = 160$

The scores provided in this way in the respective "cells" of the design are shown in Table 4.

The analysis in terms of "expectancies" proceeds as shown in Table 5. The F test shows that only the first of these sums-of-squares is significant. The data then clearly indicate that it is (b), i.e., the introversion level, which has been given the significantly large score. That is, our landlady operates in such a way as to suggest that, in terms of Jung's theory, she is introvertive. None of the other effects is significant in her case.

---

[2] This consists of arraying the statements of the sample in an order from those most characteristic of her, *in her own view*, to those least characteristic. The statements are typed on cards, one statement to a card, and the subject first reads them through in order to grasp their import. They are then shuffled, and she proceeds to the "Q sort," as we call it: usually the cards are first divided roughly into three piles, one for those that characterize her positively, and one for those that could scarcely do so under any conditions (she believes), with the doubtful or neutral ones in between. The three piles are then further teased apart, working from the two extremes, until she has provided the required forced frequency distribution. Thus 8 of the cards, which she decides are most characteristic of all gain 8 marks, the next 11 most characteristic, 7 marks . . . and so on. For ease of application we rarely employ as many items as 160 for a single Q sort. In the present case the 160 was divided into two samples of 80 each, the Q sorting being repeated for each in turn and the results combined. Sometimes a final re-sort is undertaken in such a case, in which the two sets of cards, previously arrayed by the subject in their descriptive order, are now placed in parallel before her, so that she can undertake such minor adjustments in position as she may deem necessary when all 160 so confront her. The "forced frequency" distribution has many practical advantages and is by no means as arbitrary as it may appear to be at first sight. We use only quasi-normal distributions, platykurtic in shape, for certain theoretical reasons, and we find it best to have the range of scores from 0 to 12, rather than 0 to 8 as in the present example.

Factors, therefore, was merely a mathematical model for the psychological theory of noesis: g represented noesis, and all degrees of s, anoesis. This was examined for its "general implications," and the consequence has been the complete disregard of the noetic theory in particular, and of all psychological theory about the abilities in general. Instead of theories about intelligence, purely logical metatheories have been promulgated, such as Burt's *Four-Factor Theory* (**3**), or Thurstone's *Multiple-Factor Theory* (**33**), or Thomson's *Sampling Theory* (**32**). We now know that the theory of noesis should have been regarded merely as a guide for dependent forms of experimenting, in which the theory is at issue for its singular propositions, much as we have suggested for Jung's theory above. But the early factorists were fascinated by the logic of interdependency analysis, with the result that psychological theory has become conspicuously absent from R methodology. Psychological attributes are still being studied for their interdependencies and "general implications," and the search still goes on for unitary and primary factors. It is significant to observe that psychological theories cannot be represented readily in structured R samples, whereas their scope in Q samples seems unlimited. Such is the measure of the difference between R and Q, which our critics have long believed to be merely two sides of the same coin.

Unlike the simple proposition P.1 above, those at P.2 to P.5 may require several variates for their solution, and not just one. The several constitute a variate-design, and one of the arts of experimenting in Q technique consists of designing suitable variates which will serve to put propositions in a testable form. A design for P.2 has already been reported (**31**): it was shown that members of a class who regarded themselves as extroverts had couplet factors with the experimenter, who regarded himself also as extrovert, whereas the members of the class who maintained that they were introverts did not provide such couplets. A design for the before-during-after-treatment sequence of therapy was employed by Hartley (**11**) which Mowrer wishes to distinguish from Q, calling it O technique. It is, instead, merely a particular Q variate design, of which hundreds already have been suggested or used.

These variates represent the *dependencies*, as do factors, which subsume a number of variates. Obviously it is possible to correlate such variates, and to factor the correlations: the variance analysis can thereupon be applied to *factor arrays* instead of to the original variates. In this sense factor analysis could be merely a subsumptive device, permitting us to replace many single variates by one or more factor variates, each of which can be examined for its dependencies along Fisherian lines. However, this is not the only function we require factor analysis to serve in Q methodology.

Meanwhile we may suppose that the typical *Q* study now recognizes (*a*) postulated independencies of a theory, whether made explicit in structured samples, or left implicit in unstructured ones, (*b*) theorems, hypotheses, or propositional sets having reference to the consequences of the theory, (*c*) variate designs which serve to put these sets to empirical test, (*d*) the variates themselves, and the factors which they subsume, representing dependent effects. Under certain conditions, (*e*) it is expected that the factors will point to underlying tendencies, and to lawfulness of a kind not asserted at the outset in the original theory. Factors can be interpreted, usually, in terms of the postulated theory. But such interpretations are not the essential objective of *Q* studies, which should lead to more abstract explanations which are not incompatible with these lower level ones, but which point to lawfulness and to the possibilities of general principles not perhaps previously anticipated. Thus, in studies of the self-notions of a person X we may reach factors α, β, γ. . . . These will be particular to X, no other person in the world having them if they are based on a *Q* sample of his own idiosyncratic self-reflections. But α, β, γ . . . may be in dynamic relationship, such as may suggest that if any one factor is altered (by therapy or the like) the others will change *pari passu*. It is at this more abstract level that important general principles are to be expected in *Q* studies in general, as in all scientific work.

However, we do not in fact need to structure all *Q* samples explicitly. When statements are taken which have reference, say, to psychoanalytic theory, the independencies are implied in the statements. The statements that we may take randomly from Jung's work on personality imply his theory, whether they are composed into a structured sample or not. In general, therefore, we may pursue a dependency form of factor analysis for unstructured samples, in so far as we may assert that any factors may be given an explanation in terms of the implied theory. In point of fact many propositions can be asserted beforehand for unstructured samples, just as if we had, in fact, a structured sample available.

Two forms of dependency analysis are employed in *Q* studies, that of variance design already introduced, and dependency factor analysis. Only brief reference can be made to the latter in this paper. It will be apparent to the statistician, however, that the same results can be reached, for a given matrix of data, by the variance analysis of balanced block *Q* arrays, and by their factor analysis. Effects *X*, *Y*, *Z* in variance design are orthogonal, and significant interactions are merely additional to these, but not postulated at the outset. Precisely the same effects can be postulated as orthogonal *factors*, and any not postulated can be "discovered" as an additional factor or factors. The permissiveness of the centroid method then permits us to factor our data, and to rotate until the same solution is reached as can be provided by the variance analysis of each *Q* array. The purpose of the latter analysis is mainly to permit us to classify variates which are alike with respect to specifiable effects, within stated fiducial limits. We employ factor analysis for this same purpose. The object of dependency factor analysis, it may be said, is to look at a matrix of data to determine what Fisherian design can fit it, whether one was postulated in the *Q* sample originally or not. Simple structure in Thurstone's sense (**33**) is merely such a design, albeit a confounded one when correlated factors are involved. But these are matters for another occasion.

We would like to add that the procedures we are describing apply to so-called subjective appraisals made by any subject X, or to objective ones, made about X by observers. Long ago an example was provided of the use of *Q* technique in the study of performance (**30**) as objectively regarded. The studies apply, no less, to any "single case": logically they are all best so begun. Generalizations concern abstract, higher order levels of interpretation, and are not matters having reference to "large numbers of cases" as such. Many of our most important *Q* studies involve only a single experimental subject, who may, in principle, be *anyone*.

Conclusion

The new methods we are discussing are essentially simple and straightforward in practice. It is rarely necessary in *Q* technique to deal with more than three or four factors and their combinations in order to analyze data, and no analysis need take more than a matter of a few days to complete. Nor need any variance analysis of data be oppressive: indeed it is rarely essential to resort to it in complex *Q* studies, since factor analysis provides the same results more economically.

The methodology has very wide applications, as we asserted for it at the outset (**25**). Self-psychology can now have its propositions directly represented (**11**, **22**). Many interesting studies in clinical psychology have already been undertaken (Hartley (**11**), Heine (**12**), Pemberton (**21**), Fiedler (**7**, **8**)), and also in socio-educational theory (Ebermann, **6**). With regard to social psychology we study attitudes now as theoretical issues, and not as "things" to be measured (**18**) which have this-or-that attributes (**17**). The projective techniques can now have their theoretical formulations represented as independencies in *Q* samples. Type psychologies can be more sensibly studied, freed from the erroneous beliefs engendered about them by the assumption of general propositions. Individual differences are seen as merely of technological interest, and in no way necessary as *postulates* to any psychological theory.

The methodology clearly favors a frank acceptance of theories in psychology. Nowhere in it is there any measurement-at-any-price, for no-one-knows-what. The proposal to regard factor analysis as a matter within the domain of dependency analysis is important; in this way factor analysis becomes little more than a complicated kind of t test, and nowhere is there any search for, or belief in, primary factors, unitaries, or the like. The methods we describe clearly open to our operational regard much that has hitherto been called "subjective": the only distinctions we can accept between what is subjective and what objective rests upon whether reliable operations are possible or not (Zilsel, **35**). Finally, the pertinency of modern logic and analytical philosophy is apparent in our formulations. The distinction between propositions examined for their "general implications," and theories investigated with respect to singular propositions is a case in point, but the quantitative principles upon which *Q* technique is based are such as logical analysis supports. It is because these matters were neglected that *R* and *Q* were so long needlessly confounded.

REFERENCES

1. BEEBE-CENTER, J. G. *The psychology of pleasantness and unpleasantness.* New York: Van Nostrand, 1932.
2. BROWNLEE, K. A. *Industrial experimentation.* London: H.M.S.O., 1949.
3. BURT, C. Correlations between persons. *Brit. J. Psychol.*, 1937, **28**, 59–96.
4. CATTELL, R. B. On the disuse and misuse of *P*, *Q*, *Q*s, and *O* techniques in clinical psychology. *J. clin. Psychol.*, 1951, **7**, 203.
5. CRUTCHFIELD, R. S. Efficient factorial design and analysis of variance illustrated in psychological experiments. *J. Psychol.*, 1938, **5**, 339–346.
6. EBERMANN, P. The application of *Q*-technique to one aspect of teaching competency. Unpublished Ph.D thesis. Univer. of Chicago, 1951.
7. FIEDLER, F. E. The concept of an ideal therapeutic relationship. *J. consult. Psychol.*, 1950, **14**, 239–249.
8. FIEDLER, F. E. A method of objective quantification of certain countertransference attitudes. *J. clin. Psychol.*, 1951, **8**, 101–107.
9. FISHER, R. A. *The design of experiments.* London: Oliver and Boyd, 1942.
10. GUILFORD, J. P., & GUILFORD, R. B. An analysis of the factors present in a typical test of introversion-extroversion. *J. abnorm. soc. Psychol.*, 1934, **28**, 377–399.
11. HARTLEY, M. A *Q*-technique study of changes in the self-concept during psychotherapy. Unpublished Ph.D. thesis, Univer. of Chicago, 1951.
12. HEINE, R. An investigation of the relationship between changes and the responsible factors as seen by clients following treatment by psychotherapists of the psychoanalytic, Adlerian, and non-directive schools. Unpublished Ph.D. thesis, Univer. of Chicago, 1950.
13. JUNG, C. G. *Psychological types.* New York: Harcourt, Brace, 1924.
14. KAUFMANN, F. *Methodology of the social sciences.* London: Oxford Univer. Press, 1944.
15. KENDALL, M. G., & BABINGTON-SMITH, B. Factor analysis. *J. roy. statist. Soc.* (B), (London), 1950, **12**, No. 1.

16. KEYNES, J. M. *A treatise on probability.* London: Macmillan, 1921.
17. KRAMER, B. Dimensions of prejudice. *J. Psychol.*, 1949, **27**, 389–451.
18. KRECH, D., & CRUTCHFIELD, R. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
19. LOEVINGER, J. Intelligence. In H. Helson (Ed.), *Theoretical foundations of psychology.* New York: Van Nostrand, 1951. Pp. 557–601.
20. PEATMAN, J. G. *Descriptive and sampling statistics.* New York: Harper, 1947.
21. PEMBERTON, W. A. Factors involved in clinical judgments made from projective and non-projective behavior samples. Unpublished Ph.D. Thesis. Univer. of Chicago, 1951.
22. ROGERS, C. *Client-centered therapy.* New York: Houghton, Mifflin, 1951.
23. SCHLICK, M. Facts and propositions. *Analysis*, 1945, **2**, 65–70.
24. SPEARMAN, C. *The nature of intelligence and the principles of cognition.* New York: Macmillan, 1925.
25. STEPHENSON, W. Technique of factor analysis. *Nature*, 1935, **136**, 297.
26. STEPHENSON, W. Correlating persons instead of tests. *Charact. & Pers.*, 1935, **4**, 17–24.
27. STEPHENSON, W. The foundations of psychometry; four factor systems. *Psychometrika*, 1936, **1**, 195–209.
28. STEPHENSON, W. A new application of correlation to averages. *Brit. J. educ. Psychol.*, 1936, **6**, 43–57.
29. STEPHENSON, W. Methodological consideration of Jung's typology. *J. ment. Sci.*, 1939, **85**, 185–205.
30. STEPHENSON, W. Two contributions to the theory of mental tests, II—A statistical regard of performance. *Brit. J. Psychol.*, 1940, **30**, 230–247.
31. STEPHENSON, W. The significance of *Q*-technique for the study of personality. In M. L. Reymert (Ed.), *Feelings and emotions: The Mooseheart symposium.* New York: McGraw-Hill, 1950. Pp. 552–570.
32. THOMSON, G. H. *The factor analysis of human ability.* London: Univer. of London Press, 1939.
33. THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.
34. WOODWORTH, R. S. *Experimental psychology.* New York: Methuen, 1939.
35. ZILSEL, E. Logical empirical analysis. *Int. Encycl. Unif. Sci.*, Vol. II, No. 8. Chicago: Univer. of Chicago Press, 1941.

Received January 6, 1952.

THE THREE BASIC FACTOR-ANALYTIC RESEARCH DESIGNS—THEIR INTERRELATIONS AND DERIVATIVES

RAYMOND B. CATTELL
University of Illinois

Factor analysis began with the correlation of tests measured on populations of persons, but other arrangements have since been stumbled upon, or deliberately thought out for special purposes, from time to time. In 1946 the present writer formulated the *covariation chart* (**9**) which integrated in a single conception the accumulation of existing usages and revealed certain new, logically-possible designs of factor analysis.

The purpose of this review of current practice is to call attention to some possible misconceptions and show new directions of practical usefulness. The first effect of the examination of logical possibilities in the covariation chart was to provoke a realization that up to the time of that analysis only a small corner of the universe of effective factor-analytic designs had become actively inhabited by researchers. The "chart" thus proffered powerful new covariation tools, especially in relation to the multivariate problems of clinical psychology, sociology, and physiology, which, except for a few recent examples in *P* and *Q* techniques, still need illustration. It is proposed here to develop those theorems into more explicit practical corollaries for experimental work and to investigate the true interrelations and precise limitations of the various techniques. For in some recent approaches, e.g., the use of *Q* technique by Stephenson (**31**), it would seem that there is some loss of perspective on methodological relationships.

BASIC REFERENTS IN COVARIATION INVESTIGATIONS

All scientific method deals with observations of covariation, but factor analysis covers that half of the methodological realm which has to do with simultaneous variation in many variables, not the univariate variation of so-called "controlled" classical experiment (**16**). In either region a single act of measurement has five essential referents or signatures, as follows:

1. A defined set of *circumstances*, time, place, etc., in which the attribute (reaction, trait, operation) is observed. In psychology this is the "stimulus situation."
2. The *attribute* itself, which is defined by an operation of observing or measuring certain things. In psychology this is the "response."
3. An *object*, usually in psychology an organism, to which the attribute is referred.

4. If the observation is to be quantified there is reference also to a *scale* or unit by which the measurement is to be rendered numerical.

5. An *observer*, or, in behavioral data, a set of observers capable of mutually confirmatory evidence.

Although these are exhaustive of the essential signatures for an act of measurement, each of the five is susceptible to some subdivision into subparameters. For example, the stimulus situation has many dimensions besides those of time and place required to define it, and therefore to define the measurement. However, we do not normally expressly define all of these, but merely give sufficient direction to fix and reproduce relevant *circumstances or conditions*. It is unfortunate for clarity that the term "test" is often regarded as defining both stimulus situation and response, whereas it defines wholly only the response measured. The definition of stimulus situation or occasion must be regarded as an additional referent, in which the test material is only a part. For example, an intelligence test still needs definition of the stimulus conditions in which it is given.

For the great majority of psychological experiments in which factor analysis is used we can reduce the essential referents from five to three, namely: circumstances, persons, and attributes, wherein the attribute is an operation of measurement which includes reference to that part of the stimulus situation (5) which remains fixed, and the circumstance (or "occasion") referent is restricted to whatever in the situation varies from occasion to occasion. This reduction to three referents is convenient for initial presentation of the main issues, but we shall include all five later.

If these three primary definers of a psychological observation are arranged as three distinct series (geometrically as axes) we get the covariation chart, as shown in Fig. 1, within which all possibilities of correlation for factor analytic work should be contained (except for the special extensions of the two remaining parameters).

Thus the commonest correlation is on a series of persons, each measured on two attributes, and representable in the chart by two parallel lines, as shown in the channel labelled "*R* technique," starting from two attributes ("tests") j_5 and j_7. Incidentally, it should be kept in mind that mathematically speaking these axes are not continuous or ordered but represent discrete series, i.e., populations of individual reference points (tests, persons, occasions) having any order in which the sampling happens to present them.

Any pair of parallel lines drawn within the parallelepiped of the covariation chart will represent a possible correlation, for it will indicate measurements of one character, made in two different forms upon a

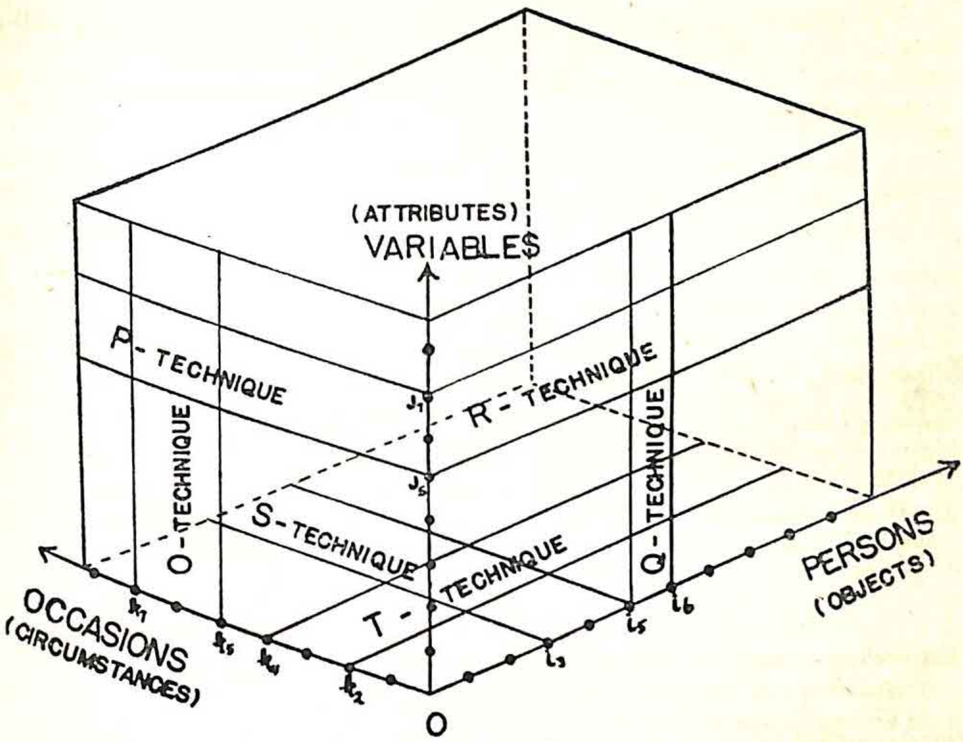


FIG. 1. THE COVARIATION CHART

series belonging to a single class. Thus in addition to the correlation of tests j_5 and j_7 for a series of people, as just illustrated for the classical *R* technique, we can draw a channel to the left (Fig. 1) which represents a correlation of j_5 and j_7 upon a series of occasions $k_o \cdots k_n$, for one person i_o.

Or again, we can take two occasions, as shown at k_2 and k_4 and correlate the series of people $i_o \cdots i_n$ on a test j_o, as labelled *T* technique. This, incidentally, is a reliability coefficient, and a whole matrix of such pairs of occasions could be factorized to find "factors in occasions (circumstances)" producing similar behavior on a test. Channels drawn in any one *plane* amount to correlatable series in which the same thing is held constant. For the present we propose to refer only to rectangular series drawn parallel to an edge, omitting the special problems of "staggered" (lead and lag) correlations, etc.

THE UTILITIES OF THE BASIC DESIGNS

It may help to fix the above six designs in mind, for the purpose of further abstract reasoning about them, if we expand briefly on the special scientific utilities of each.

1. *R technique*, factoring attributes on populations of persons, is the daily bread of the psychometrist and needs no description. Its factors are *common traits*, applicable to measuring individual differences of persons and having a meaning rooted in the behavioral relations of *a whole population and its environment.* If simple structure is applied, to give natural, functional, and not merely mathematical factors, the factors will tend to be invariant in loading pattern, differing from sample to sample principally in slight changes of obliquity of the factors (**35**).

2. *P technique*, correlating attributes within one person, is, by contrast, the agent for discovering *unique traits* and particularly for that unravelling of connections of dynamic traits and symptoms for which the clinician has hitherto had nothing more positive than free association (**19**, **20**). It is an ideal method for determining, along with other dynamic structures, the structure of the self sentiment, as shown in two recent studies (**19**, **22**). In two other studies it has shown a potency unequalled by any other method but fully controlled physiological experiment for revealing the connections of interest in psychosomatics (**20**, **38**). Its value in sociology, economics, and social psychology has been shown in two studies factorizing longitudinal series for single communities or nations, thus introducing a more positive and precise calculus in relation to historical influences and trends (**17**, **18**). By this method both the factor structure and the quantification of the factors are unique to the individual social or animal organism, the scaling thus having to be ipsative (**8**), whereas in *R* technique the scaling is normative and the uniqueness of the individual is only a uniqueness of pattern in *common* dimensions.

For the above reasons the factors obtained by *P* technique have no *necessary* mathematical relation to those obtained by *R* technique. The half dozen pioneer studies so far published suggest, however, that a *scientific* relation will exist in that the structural patterns of the individual unique factors will tend to scatter about the loading pattern of the corresponding *R* technique common factor. The *R* and *P* factors may also differ in total variance contribution. For example the surgency-desurgency factor of elation-depression seems to stand higher in factors of intra-individual variance than in inter-individual variance (**13**, **19**, **20**, **38**).

3. *Q technique* has its chief use as a classificatory device for finding the subpopulations in a nonhomogeneous population (like Lazarsfeld's "latent structure analysis" or latent subgroup analysis), and for the special purpose of quantifying the extent to which an individual may be regarded as belonging to certain species and subspecies. Whether this can be considered a process of defining types depends upon the meaning one assigns to the term "type."[1] In social psychology it has value in picking out roles, i.e., common patterns of social re-

---

[1] I have suggested elsewhere (**8**) that "type" is being used for four distinguishable concepts: (*a*) continuous types, e.g., a tall and a short type of man, where a whole factor pattern varies in level of elements continuously from extreme to extreme; (*b*) discontinuous types, e.g., man and pigmy, where the pattern is essentially the same as in continuous types but the measurements are bimodal or discontinuous; (*c*) continuous species types, where the pattern itself differs, but in a continuous way, e.g., business men and artists; and (*d*) discontinuous species types, e.g., dogs and ducks, where the pattern is distinct and discontinuous as to distribution in nature. Some uses of *Q* technique to find types slur these differences.

sponse showed by many persons; but this can be dealt with also by *S* technique described below. Strictly, *Q* technique is a factor analytic design and as such should proceed *beyond* the mere examination of a matrix for "type" *clusters* to the further derivation of abstract *factors*. If it does this it may have the additional utility of providing a new, independent avenue to discovering the universal factors of *R* technique, and an avenue which may be more convenient than *R* technique in some special circumstances. This transposability of *R* and *Q* technique results is denied by Stephenson, though Sir Cyril Burt claims to have demonstrated it (6) and some statisticians agree with this demonstration.

4. *O technique* is perhaps the more important of the two *occasion-correlating* techniques. It tests one person on a whole series of tests for two occasions (k_5 and k_7 in Fig. 1) and determines the similarity of the total person (or, at least, as much of him as is included in the series of tests) on two occasions. The most obvious use for the factorization of a matrix from such pairings of occasions is for investigating multiple personality (9, p. 99) and the change of the self-structure under psychotherapy. The nature of the factors has to be inferred from the attribute pattern for those occasions which are most highly loaded in the given factor. This approach also lends itself to analysis of stimulus situations, for the nature of the circumstances in each occasion can be manipulated, or at least recorded, and the factors among occasions will thus show how situations group themselves in regard to their effects on the total personality. Some alleged instances of *Q* technique are really *O* technique, and would be more correctly evaluated if this were explicitly recognized.

5. *T technique* may be remembered by the mnemonic that it is *test-retest* factorization. Like *O* technique it correlates occasions, but it does so, as the covariation chart above shows, by repeating the same test for a matrix of different occasions on a population of persons. It is thus a factorization of reliability coefficients, but naturally the significance of the design would hinge upon recording the particular *circumstances* of each occasion of re-administration, as in *O* technique. The factors among such defined occasions would again be factors in the general stimulus situation as judged by effects on the population. But they would differ from *O* technique in being factors with respect to some single response, e.g., an opinion poll on a definite issue, instead of to a sample of the responses of the total personality, and in applying to a population instead of a single person. Because of this analysis of circumstances according to the effects they produce on the distribution of response with a population, *T* technique might be called a "social climate" thermometer and is likely to find its most valuable application in quantifying phases of business cycles and in getting more historical meaning out of opinion polls than the present poverty of statistical methods permits.

6. *S technique* may be remembered by the mnemonic that its principal immediate use is to detect and define *social roles*. It correlates two persons on their reactions to a single stimulus (test) on a series of occasions. (A role is defined as a pattern of responses to different occasions which is modal among individual patterns, i.e., it is a cluster or factor among people in responses to social occasions.) The loading of a person in the factor would show the extent to which he is successfully assuming the role. *S* technique also has promise in determining the internal structure of a group or institution; for example, by correlating the responses of members of a family over many occasions on their response to a single issue it would indicate the functional subgroups within the family. Be-

cause of this capacity to reveal the degree to which individuals belong to groups characterized by homogeneity of functioning in regard to a series of historical events, S technique is *par excellence* the method for social psychology.

THE INTERRELATIONS OF THE SIX TECHNIQUES

In respect to inherent, systematic characteristics there are in fact two main kinds of relationships to be considered among the six techniques, and these are again most readily perceived by reference to the covariation chart itself (Fig. 1).

First, we observe that two techniques may depend on correlations which lie in the same face of the parallelepiped, which means that their correlation matrices share the same entries, i.e., they work upon the very same data but correlate it in different arrangements. This gives us the three pairs of groupings R-Q, T-S and P-O. Thus the first pair, R and Q, share a matrix of test measures upon people; P and O share a matrix of test measures upon occasions, and so on. (Or alternatively, they hold constant the occasion, the person, etc.)

Second, any two techniques may share an edge of the model, i.e., the parallels representing their correlations hinge on the same edge. This means that they are seeking relations among the same variables but upon different modalities of population. The correlations of the same variables in the two techniques have different meaning because they deal with different influences and a different source of variation, yet because they are the same variables both techniques are required as supplementary statements of the nature of the factors in the variables. This groups the six techniques in the three pairs: R-P, Q-S and O-T.

Now the first of the above groupings distinguishes *among* the pairs on the grounds of modality of data, i.e., the members of a pair work on common referents, while *within* the pairs the distinction is one between complementary statistical processes. Indeed, we shall argue below that since the complementary processes rearrange the same data their results must in general be statistically transposable and that the members of a pair are really equivalent. The researcher's interest in bringing the two methods of one pair to bear upon any problem is therefore to obtain statistical confirmation of the same scientific relations, and to check the soundness of the statistical part of his research by proceeding from data to conclusions through two routes.

By contrast, the pairings according to the second principle guide him to possibilities of independent experimental or general scientific confirmations. Thus R and P techniques have been used with precisely the same batteries of personality measures (**9**, **12**, **19**, **20**) but, respectively, upon a whole population at once and upon a series of individual studies of clinical cases. It has been found that some of the dynamic unities revealed by the one are also revealed by the other (**11**, **14**, **17**, **22**). Similarly in the *S-Q*, and *O-T* pairs of techniques we may in general expect that connections between variables in one population will be borne out if the functional unity of the trait is universal by their connections also in other realms of variation.

For clarity of further reference we shall call the three pairings of the first type *common matrix techniques* (yielding *transposable factorizations*) and those of the second *common variable techniques* (yielding independent but *common function factorizations*). Thus *R* and *Q* are common matrix techniques because both begin, or can begin, with the same score matrix.[2] In the first case one correlates the columns (tests) and in the second the rows (people). It is the writer's opinion that psychological research today stands most in need of appreciation of the gains possible from a two-handed use of the *common variable* techniques on any one scientific problem in mutually checking studies. But since some current bandwagons of enthusiasm have stressed the common matrix designs to excess, it is necessary to digress into considerations which will help perspective.

The misperceptions that, in the writer's opinion, are most prevalent are (*a*) those shown in attempts to obtain by *Q* technique what is better obtainable by *P*, *O* and even *R* technique, and (*b*) in a denial of any statistical relationship between the results within each pair of common matrix extractions, e.g., of *R* and *Q* techniques. As to the latter we should note that the original development of *Q* technique (though not under that label) by Sir Cyril Burt in 1912 (**2**), 1917 (**3**), 1931 (**4**), and 1933 (**7**) occurred expressly with the intention of arriving at the ability factors simultaneously hypothesized in *R*-technique studies. In Stephenson's 1935 article (**29**), "Correlating Persons Instead of Tests,"

---

[2] It seems to have been misunderstood by Stephenson that this label refers to common form of matrix, not necessarily to such furnishing as hypotheses, or the subjectivity or objectivity of data, number of entries, distribution of variables, etc. In designing a test-person matrix for *R* technique the number of persons, *N*, will usually be larger than the number of variables, *r*, while for *Q* technique the number of available persons, *n*, is usually small and the range of variables, *R*, is large. Naturally this difference of emphasis will also be associated with differences in sampling of variables and persons and the choice of hypotheses. Such differences for that matter will also occur within examples of *R* technique itself, according to the particular design and purpose of the experiment. But the *essential* fact is that both use a test-person matrix, and if one wishes, as in a recent study by the author using 100 tests on 100 persons, the *identical* matrix can be used, first read upright and then sideways. Stephenson's quibble would amount to saying that since some books are taller than they are wide and other the converse, it is a great mistake to recognize that they are all books!

Burt's design was acclaimed with new enthusiasm but lost some of these earlier perspectives and insights. Incidentally, Burt (**5**) credits the germinal idea to Stern (**32**) who in 1911 wrote of "horizontal correlation" and "vertical correlation," though without the superstructure of factor analysis, and he points out also an early use of Q technique by Beebe-Center in 1933 (**1**). Incidentally, Stephenson titled his basic article "The Inverted Factor Technique" (**30**), i.e., the inverse or obverse of R technique. However, statistically-minded psychologists have preferred *transposed* factor analysis, for, to be algebraically exact, Q technique factorizes the *transpose* of the R-technique matrix.

Accordingly, here and elsewhere the writer follows the convention of calling Q transposed R technique, while S is similarly transposed T technique and O is transposed P technique. If we are correct in supposing that the transposition is a purely statistical operation—a matter discussed in the next section—there are basically only three independent factor-analytic experimental designs, namely, R, T, and P techniques. These alone yield factors in statistically independent, unbridgeable measurement systems, and are the pillars between which bridges of paired scientific inference and relation can be built.

THE MEANINGS AND UTILITIES OF THE COMMON MATRIX TRANSPOSES

It is necessary to stress in the penultimate sentence above that we speak of experimental designs not experiments. For an R- and a Q-technique experiment using samples of tests and persons from the same populations would be independent experiments, but only in the sense that two R-technique experiments on related data would be independent.

The belief of some users of Q technique that it is fundamentally different from its transpose technique—R—and, indeed a method *sui generis*, has so far been most exhaustively statistically examined and refuted by Sir Cyril Burt (**5**). Saunders' work on "direct factorization of score matrices" (**27**) supplies an incidental proof of their interdependence from a new angle. In the writer's experience professional statisticians take the position that there is no doubt about the transposability of factors from a double-centred score matrix though there may be doubt about the exact relation under other and special conditions. A paper by Hicks[3] and the present writer aims to deal with these statistical issues in more detail.

---

[3] HICKS, V. E., & CATTELL, R. B. The problems of transposing factor solutions in common matrix factorizations, illustrated by an example in R and Q techniques. In preparation.

Here it is possible to examine the relations of transposes only in an introductory and illustrative fashion, as far as method is concerned, proceeding mainly from first principles in the use of the correlation coefficient. However, we shall do so comprehensively from the standpoint of experimental design, under five headings as follows.

1. *Transposability of variance findings, with partial loss of variance information.*[4] It is relatively easy to see that R and Q (or P and O, T and S) techniques normally (i.e., without double centering) have the completeness of their transposability slightly restricted by some inevitable mutual losses of information. The losses which then occur are (*a*) of the variance of the first factor (or in some conditions the first two) and (*b*) of the specific factors.

The first kind of loss can be most readily perceived by a concrete example, say that of correlations upon a matrix of physical measurements, e.g., leg length, size of shoe, sitting height, weight, for a population of men. If we correlate such "tests" the first R-technique centroid factor is likely (and indeed known) to be a "general body size factor," since the man with greatest stature is likely to have the largest boots and the greatest weight. But if we correlate persons, nothing corresponding to this is discovered, for a small man and a large man of similar proportions will correlate perfectly. Size is overlooked, because the correlation coefficient "responds" to similar profiles, regardless of level.[5] The correlation coefficient in effect behaves as if it has scaled the raw scores to standard scores, as far as the columns being correlated are concerned, reducing them to the same mean and the same sigma.

Reciprocally, R technique loses a factor through virtually bringing the *tests* to the same means and sigmas. The fact that all men are taller than they are wide is overlooked by the correlation of height and width, but not in the corresponding (raw score) Q-technique procedure. The first centroid in Q technique is largely due to the correlations arising from the fact that all people resemble each other to some extent, e.g., they are all taller than they are wide. This massive factor in the matrix may be called the "common species" factor, for it defines the common pattern of the population and the individuals's endowment therein represents the extent to which he resembles the species.

The relative importance of these reciprocal losses will be argued below, but let us at least notice here that most psychologists, on common-sense grounds, are inclined in Q technique to *standardize* the rows of tests (they are rows in the transposed matrix) before correlating the columns, for they point out that the raw score in which a test is couched is arbitrary and irrelevant. For example, if, in the above illustration, body weight is measured in ounces, this variable will be highest for every individual, whereas length of nose (in feet) will be lowest for everyone. The important thing, it will be argued, is whether the person's nose is large for a population of noses.

---

[4] This section is concerned with the difficulties referred to in Stephenson's comment that "it has long been difficult for everyone except myself to accept the proposition that R and Q are obviously completely different." (Contribution to a symposium held at the meeting of the Midwestern Psychological Ass., Chicago, Ill., April 27, 1951.)

[5] The r_p "coefficient of pattern similarity" (18) takes account of difference of level as well as of shape. If we had some experience with its use in matrices for factor analysis it might transpire that an improvement of Q technique would be obtainable through substituting r_p for r.

Such a standardization of rows when one is correlating columns of persons, or, in R technique, a standardization of rows of labelled persons when one is correlating tests, takes the *species factor* out of Q technique and the *size factor* out of R technique, so that both now lack the same two "potential" factors and, for the rest, have factor solutions which could be truly transposable. The general statement that it is possible to use either technique as a gateway to the same end result, therefore, needs the modifying clause that with usual scaling procedures one (or two) general factors are lost in the process.

The relative importance of the factors reciprocally lost in the transposes remains to be evaluated. For most purposes the "species" factor lost in R technique in the present example is less important than the "size" factor lost in Q technique, for we know the species with which we are dealing and are most interested in individual difference within the species, having nothing to do with species character. Furthermore, though Q technique necessarily loses a second factor, for reasons given below, there is really no need to lop off the second factor in R technique, so that the latter in fact loses no information of interest and value in the predictions with which psychology is chiefly concerned.

The first potential Q-technique factor—that of general size in our present example—is lost inevitably. The second is pruned away by that use of standard scores which most psychometrists rightly demand in this situation, wherein the absolute scores of variegated tests are meaningless. But in R technique the corresponding pruning—by standardizing persons instead of tests—is not only unnecessary but objectionable. Bringing different tests to the same mean and standard deviation, despite wide differences in raw scores, makes sense because there is no meaning to such a statement as "Mankind as a whole is higher in mechanical aptitude than intelligence."

Let us next briefly examine the situation regarding specific factors, i.e., those peculiar in R to a test and in Q to a person. In the first case a group of people covary in their performance on one test only, and in the second one person has a covariation on several tests not shown by anyone else. It would seem that these are inevitably lost in the transformation. Simultaneous variation on tests and persons must exist for transposition to occur. Light is thrown on this by Saunders' K-way scale analysis (**27**), proceeding directly from the score matrix.

The mutual losses in R- and Q-technique transformations should be further considered by the specialized reader in the general perspective regarding losses of information when factoring (*a*) covariances, (*b*) score matrices, and (*c*) correlations, or, indeed, in the perspective of statistical abstracting generally.

2. *Transposes examined in terms of experimental convenience.* So far the discussion has proceeded as if we dealt with a roughly square matrix, correlated on a set of columns for R technique and turned on its side and correlated on what were the rows for Q technique. But the labor of dealing with as many variables as one needs to have people, for a reliable sampling of the population, is too great and, in practice, the score matrix has always been trimmed to oblong form. Therein the short side—that representing the things to be correlated—consists of tests in R technique and people in Q technique. The resulting failure of symmetry adds further differences between the two techniques though they are differences of degree, not of kind. What we need to discuss under this heading concerns the claim by some clinicians that the same result may be obtained with less labor by Q technique. Actually, if we preserve statistical equivalence (equivalence of reliability), the decrease in subjects is balanced by an increase in tests and the number of subject-hours of testing time (or the cost of subjects) remains exactly the same. Since much testing can be group testing, however, there is more saving of the experimenter's time through R technique.

The respective availability of time and subjects, therefore, may well dictate the choice as far as this very practical consideration is involved. But there are also questions of "convenience" from the standpoint of reconnoitering hypotheses and the strategy of research. These will be discussed below in relation to analysis of variance and testing of hypotheses.

3. *Transposes in terms of generalizability of findings.* It will be evident that despite the complete *logical* symmetry existing between the transposes we have to admit at several points qualifying conditions on the reciprocity, owing to different inherent properties in the "equivalent" series, etc. This needs to be kept in mind most steadfastly in regard to the sampling problems and the resulting restrictions on generalization. For the sampling of tests, with our existing knowledge and concepts, cannot be handled in the same way as the sampling of *people.*

The argument that restricted sampling of people is more dangerous than restricted sampling of tests does not seem to have been seriously appreciated everywhere, though it has been duly regarded in Burt's original writings on Q technique. Stephenson's position[6] appears to be that if he can establish certain patterns of relation among people in a small group he need not be concerned with their wider generalizability. This betrays a fundamentally different assumption about the aims of science, substituting a descriptive and revelatory (one might almost say artistic) goal for the usual one of explanation and generalized prediction.

---

[6] His assertion that even a single case (a crucial experiment) can prove or disprove a general scientific law seems to the writer to confuse the distinction in scientific method between controlled experiment, reaching its extreme expression in some experiments in physics, where all irrelevant variance is controlled, and statistical analysis of variance *in situ* and uncontrolled, as it has to operate in most social science research (**16**). It may also confuse description and explanation.

If only half a dozen persons are taken in Q technique the results can be generalized to the parent population only with the high degree of guesswork which that very small sample permits. The fact that tests have been multiplied does not save one from the laws of sampling inference where people are concerned. It should be clear, however, that this argument does not deny that the same general life processes and logic actually operate in a small group, or a single person, as operate in the parent group. It only denies that such methods of investigating small groups or individuals, without regard to statistical needs, *are capable of finding them*. Indeed, the boot is on the other foot, for it is the present misusers of Q technique who deny that *universal* processes, particularized to the situation, explain the behavior in the small group or individual; for they claim that the processes which *can* explain are understandable without any experience of their manifestations elsewhere. To which one can only reply with the logician that the meaning of "A" requires knowledge of "not-A," or with the poet "What know they of England, who only England know?"

Naturally we should ask if an equivalent objection applies to the reciprocal situation of the small number of tests used in R technique. This is where statistical symmetry is no longer a scientific or epistemological symmetry. *The test population, except for the personality sphere concept* (**9**), *does not have the qualities of a biological species population.* The experimenter, indeed, usually does not even make any attempt to sample the test universe. He establishes factors in a given area, staked out by defined and preserved tests, and then, in later experiments, he goes on to extend the area. He does not say "These are all the personality factors in existence," but "These are the factors *only in the area I have so far staked out* with the defined tests, *but* they are true of *all* people." The equivalent procedures for Q technique are by no means so dependable, e.g., one cannot preserve people as "markers," as one does specific tests.

There is thus in any pair of transposed common matrix techniques a tendency for one of them to rest on a surer sampling foundation than the other and to permit wider generalization with the same amount of work.

4. *Comparison of transposes in terms of reproduction and interpretation of results.* As indicated incidentally above, the interpretation of factors and the preservation and communication of defining data are undoubtedly rather less efficient in Q than R technique. Where R-

technique factors require the filing of two or three tests and a note on the population sample, Q technique requires that we file a very lengthy profile of performances for the highly loaded individuals—or file the individual himself!

Interpretation of the factor in R technique requires that we contrast the nature of the highly- and the zero-loaded test. This is at times a fairly difficult process of abstraction and inference, but it does not present the double act of abstraction required by Q technique, where we compare a whole series of test scores for the person of high loading with another series for the individual of low loading.

To examine the living person himself is no easy alternative. For although a person may be, say, highly loaded in the factor of intelligence, his every act presents intelligence inextricably interwoven with every other ability, emotional, and temperamental factor in his nature; whence recognition of what "characterizes" him is not easy.

5. *Simple structure and factor invariance in the transposes.* Assuming that invariant, stable factors have so far been producible only by rotation for simple structure, the problem of how effectively invariance can be obtained in the transposes reduces to an examination of what simple structure means in each. The scientific assumption that most factors will operate significantly only in a limited number of any widely chosen battery of test performances, on which R technique rotation rests, is logically acceptable and empirically verified. In Q technique it is logically less acceptable that some *persons* will be *wholly* devoid of a factor operating in others.

Discussion of rotation requires the statement of our general position on Q technique, which is that its true contribution (apart from occasional convenience in approaching R-technique factors through a transpose) is specifically with *nonhomogeneous populations.* Then Q technique or the method of latent (structure) subgroup analysis, carried perhaps only to the point of finding correlation clusters, can be used to find the separate "species types" or homogeneous subgroups upon which R technique can be most profitably employed. Thus in a group composed of, say, trained artists, soldiers, and philosophy students, a simple structure rotation could be expected, *provided the variables deal only with those aspects of personality which are much affected by occupation.*

Apart from such nonhomogeneity one would expect the distribution of Q factor loadings to have a substantial proportion of zeros only because of the effects of normal distribution of loadings.[7] Unfortunately there seem to be no *Q*-technique studies substantial enough to have been examined for empirical evidence as to an invariant rotation position.

---

[7] The frequency of essentially zero loadings in a well-sampled R-technique battery is typically around 60–70 per cent with respect to any one factor (**10**, **12**, **14**), which should be distinguishable from the smaller percentage of apparently zero loadings resulting from the application of a smooth normal distribution curve to the given factor loading distribution.

If no universally acceptable method of ensuring invariance in *Q*- and *S*-technique factorizations should be found, their whole utility falls to the ground and they cannot be used as alternative avenues to the general universe of scientifically negotiable results provided by their transposes, *R* and *T* techniques.

DERIVATIVE FORMS RELATED TO SCALING AND SUBJECTIVITY-OBJECTIVITY OF EVIDENCE

The discussion has so far dealt with the patterns of correlation possible by series arrangeable among the three chief signature characters of a measurement, but it is evident that, if the two remaining signatures are also considered, fairly numerous additional combinations are possible. For practical purposes only two or three possible categories actually make sense in these series, and the total of six designs so far considered is in the end only doubled or quadrupled. However, it is most important to deal with these further modifications because in some recent writings they have been confused with the primary differences, e.g., *Q* technique has been said to differ in terms of scaling (**31**).

Scaling differences rarely affect factor-analytic conclusions in themselves. But let us consider the principal alternatives, which are: raw scores, normative scores (ipsative (**8**) within a single person, as in *P* technique), scales of the Guttman type, logarithmic scales, percentiles and raw scores forced to a normal distribution. The only modifications relevant to factor analysis are the forcing of arbitrary raw scores to a normal distribution, which Thurstone (**36**) has shown to give clearer factor structure, and the change to normative scores across the *arrays* (rows or columns) correlated, which we have shown above to eliminate a factor. Stephenson (**31**) speaks of four "foundations of psychometry," meaning standardizing rows, standardizing columns first and then rows, and so on. As indicated, the two last are superfluous, for the use of r automatically produces standardization one way. The real alternatives are not to standardize at all or to standardize across the arrays being correlated. Much discussion at cross purposes might be saved by a convention of writing these as Q and Q_1 or something similar.

The distinctions connected with the fifth signature of a measurement (nature of the observer) are, however, more fundamental than those of the fourth (scaling). Some workers in personality factorization have adopted the convention of writing *BR*, *Q*, or *OT* alongside the factors (**9, 10, 12, 13, 21, 22, 35, 38**) to indicate respectively whether they rest on behavior rating *in situ*, on questionnaire data, or on objective (nonintrospective) behavior data in the special situation of a test. From numerous scientific standpoints the results from these three sources need to be differently regarded and differently treated, but from the broader viewpoint of the basic character of the observations the difference between behavior *in situ* and behavior in a specially devised test is irrelevant and we have only the distinction between behavioral data and subjective, introspective, questionnaire data (the latter when treated at face value).

Broadly speaking the present writer is in agreement with such behaviorists as Spence (**28**) that introspective data cannot be truly integrated into scientific psychology. The extensive use made of such "data" in some recent Q-technique work is consequently abortive. But this evaluation must be qualified by the recognition that (a) the "mental interiors" of factors (**12**, **14**, **21**) are of interest even if we cannot integrate them and they remain mere epiphenomena, (b) in reconnoitering stages of research and with subjects of blameless motivation they are a rough guide for later objective research, and (c) the objection disappears if the introspective reports are treated only as behavior. For the scientific restriction is not on using the results of introspection, or on questionnaires, or on verbal responses, but on taking as data that cannot be witnessed by any second observer. In other words, in factor analysis this amounts to using correlations where we cannot get a reliability coefficient calculated between data obtained by two distinct observers.[8] *Verbal responses are acceptable if the experimenter refrains from accepting the conventional face-value of the words as symbols* and independently establishes their relation to other behavior, which could best be guaranteed by getting them in Urdu or Swaheli or some other language he does not understand.

In the more systematic R-technique studies the meaning of questionnaire (or R_s) factors has been established by their correlation with behavioral factors (**21**), but in recent Q-technique studies of the self-concept the temptation has prevailed to accept introspection instead of inferring the self-evaluation from behavior. Since, as Freud and sundry philosophers have observed, language is not only a limited but also a systematically distorted form of behavior, a correlation expressing a person's resemblance to another (Q technique) or a person's resemblance to himself on another occasion (O technique)—even when language is treated strictly as behavior—has all the unreliability and meaninglessness of working with a quite inadequate sampling of the relevant parent "population of variables." Actually the personality studies of Stephenson (29) and Rogers (26), as argued in a recent article to clinicians (15), are Q_s- and O_s-technique studies, not integratable with behavioral factor analysis. For though statistical treatments of introspection are interesting, as agreed above, they relate to behavior only as a non-Euclidean to a Euclidean geometry.

---

[8] The observers must be peers in objectivity. If even one is introspecting, the whole belongs to factorization of the subjective.

Burt has pointed out that the general negotiability of Stephenson's Q_s technique is additionally confused by the almost mystical instruction on "significance" which every subject, bright or dull, is required to understand in order to participate in the experiment. When subjects are asked to rate traits according to their "significance" to their personalities, one may surely expect that the differing subjective perceptions of this instruction will add an additional dimension of error to the subjectivity already inherent in the replies. R_s technique at least avoids this difficulty. The best statistical meaning that can be given to "significance" is that the subject is being asked to rate the various aspects of his behavior or feelings on a continuum of "eccentricity" or deviation of those kinds of his behavior relative to other kinds—all in relation to an implied norm. In R_s one asks him the difficult question of how deviant he is in say "liking Rembrandt" or in "computational skill" relative to the general population, but in Q_s we pile on a demand for the further judgment of how much more deviant he is in liking Rembrandt than in his computational skill!

THE CONTEXT OF FACTOR TECHNIQUES IN GENERAL SCIENTIFIC METHOD

This concluding section proposes briefly the setting of current factor analytic designs in statistical and scientific method.

As stated at the opening of this article, the observations of covariation which suggest or test scientific hypotheses may be either (a) univariate or (b) multivariate. In the former the variable observed is the *dependent variable* changing in response to controlled changes in the independent variable. In the latter the variations of all are observed simultaneously with no temporal or other priority for any one of them. The former has usually been associated with actual experimental manipulation whereas the latter is of value in psychology and sociology, where some of the most important things cannot be controlled in the laboratory but have to be examined *in situ*. The former *tries* to keep everything constant but the pair of variables concerned, and what it cannot control it calls *error*. The latter controls little or nothing but keeps track of what actually happens to many variables which would be unknown sources of error in the former. (It still leaves some unknown variance in specific factors.)

The statistics of the former are mainly concerned with significances of differences of means and analysis of variance, while the latter uses multiple correlation, the discriminant function and, principally, factor analysis. For brevity we may discuss factor analysis and analysis of variance as representatives of the two halves of statistical method which need to be related if one is to see more clearly how factor analysis functions. Incidentally, the present writer (**16**) has proposed a hybrid of these two in which some degree of experimental manipulation and control of variance is combined with factor analysis.

The statistical differences of factor analysis and analysis of variance are embedded in profound differences of strategy in scientific method. In the latter the experimenter assumes he can test (or produce) his hypothesis by a single variable (or pair), i.e., that he knows already which variable is most important. In the former he realizes that in a bewildering array of variables the significant variables (factors, latent variables, variables of greatest influence in the field) have still to be found. What is more, he recognizes that frequently the concept in a hypothesis cannot be operationally defined in a unique way by one variable. It is a paradox of this methodological contrast that whereas in analysis of variance the experimenter generally assumes more insight into the relevant hypotheses, he applies a less severe test to them (trusting to a one-variable manifestation instead of a pattern) and gets the information only that "a significant relationship exists" instead of a quantitative statement of the degree of relationship, as in correlation and factor analysis.

Looking at psychological research of the past thirty years one cannot avoid the conclusion that factor analysis should more frequently have preceded controlled experiment and analysis of variance. For example, there have been countless experiments in learning relating dependent variables to degree of hunger, and in personality in relating life adjustments, etc., to degree of extraversion. The experimenter has generally been content to represent each by one variable (hunger by hours of deprivation, extraversion by some arbitrary rating from Jung). A factor analysis should first have been performed to see if a factor pattern could be found corresponding to each of these concepts and to find the pattern of weighted variables most accurately estimating each in further experimental work.

Since the object of juxtaposing this article with Stephenson's in the present issue is to bring out differences of viewpoint, I must turn to what seem to me misrepresentations on his part of the above discussed roles of factor analysis in relation to analyses of variance and scientific method generally. Stephenson's proposal to relate analyses of variance to *Q* technique does not seem to me to add anything not already recognized in these methods. It does not present the novelty of experimental design found in the more radical hybrid proposed above (**16**). As an expanded and more brilliant treatment of the sketch just presented of the relationship of the two methods, the reader is referred to a recent article by Burt (**6**) who points out that analysis of variance may be used to demonstrate but not discover factors,[9] to test their significance but not their nature (pattern).

But the most misleading assumption is that *Q* technique has a monopoly (presumably among the six factor-analytic designs) of "hypothetico-deductive direction . . . singular propositions . . . transitory postulates"—in short of hypothesis testing! Stephenson's assertion that hypothesis testing "can find [no] place in *R* technique"[10] can be refuted without going further afield than the work of my immediate colleagues, where the design and choice of variables in objective personality tests (**10**) and the investigation of ego defense mechanisms (**37**) were guided by several highly explicit, testable hypotheses. Or again, in the first experiment (**19**) relating *R* and *P* techniques it was hypothesized that the same factor structure would be found in any one person as in the general population, and every single variable and condition was carefully chosen in relation to this theory and five specific subhypotheses. Or, yet again, we may note Stephenson's still more naive assertion that *Q* technique enables us "for the first time in history to operate with testable assumptions about the self."[10] It has evidently escaped his notice that *R*-technique studies, with objectively measured attitude-interest variables have already (**13, 14**) been designed on a set of explicit hypotheses about the structure of the self-sentiment and have emerged with contingent proof of the most important of these.

All variants on the three basic factor-analytic designs—*P*, *R* and *T* techniques—share the power of factor analysis to *produce* hypotheses more readily than univariate methods and to *test* them more searchingly. The regularity of covariance, the lawful behavior, observed in a pair of variables cannot suggest such overdetermined hypotheses as can the emergence of a factor.[11] Re-entering a factor-analytic experiment with the hypotheses corresponding to the factors first found permits a more searching test of them, because if a given hypothesis is to be verified, not one variable but a whole pattern of variables has to behave in the manner predicted by it.

---

[9] A group of tests (or a set of persons) sharing high endowment in the same factor will have (standard) scores with significantly less within-group variance than between-group variance with relation to other groups also largely made of one-factor measures.

[10] Quoted from Stephenson's paper which was read at a symposium held at the meeting of the Midwestern Psychological Ass., Chicago, Ill., April 27, 1951.

It may be questioned, however, whether Q technique as used by Stephenson and Rogers (Q sort) is getting the full advantage from the hypothesis-producing power of factor analysis. When Rogers starts with such a hypothesis as "where the self-concept is formed entirely from the evaluation of others the individual will at some point face internal conflict" (**26**), one is evidently dealing with something resulting from a casual glance of the naked eye rather than a precision concept gained by the factor-analytic microscope.[12] And when Stephenson proposes to "take the guesswork out of factor analysis" by using only variables from a preconceived theory—gained from an inward eye, or Jung, or Aristotle—he is neglecting the more precise hypothesis formation possible from first seeing what lawful relations factor analysis will show in a truly varied and comprehensive array of variables. To most American psychologists this Old World subjectivity amounts to putting guesswork *into* factor analysis, not taking it out. In too many of these discussions we are dealing with an inverted semantics, not merely an inverted factor technique.

SUMMARY

1. There are six primary factor-analytic experimental designs, defined as O, P, Q, R, S and T techniques.

2. As far as statistical independence is concerned, these reduce to three independent *common matrix* pairs: $R(Q)$ technique, $P(O)$ technique, and $T(S)$ technique. In each pair, one is the transposed technique of the other, and, with attention to required conditions, can be

---

[11] This is the logical point at which to make reference to the fatuous, but, one hopes, now less fashionable, slogan that "one only gets out of a factor analytic design what one puts into it." This is actually a truism of all experiment—one cannot find relations between variables one has not used in the experiment. But the hypothesis about structure with which one emerges may be entirely different from those with which one entered the experiment, e.g., one may find seven factors where it was asserted there were only two.

[12] How many people exist whose self-concept is formed *entirely* from others? And is it not safe to say that *everyone* faces internal conflict? One would want to know what factorial proof has been given of the unitary dynamic nature of this alleged entity the "self concept." And, of course, the *degrees* of conflict associated with the *extent* of its formation by associates could be most readily investigated by R technique.

used as a second avenue to the discovery of the same factors (though a fraction of the group and all the specific factor information are lost in the process).

3. The six experimental designs can also alternatively be reduced to three *common variable* (and therefore common loading pattern) pairs. Within these there is no *statistical* transpose equivalence as in common matrix pairs, but a scientific *experimental* equivalence in that the same functional entities may be tested in two different contexts of manifestation.

4. Since any measurement has five basic referents or signatures which particularize it, the above six combinations among three of the referents (forming the covariation chart) can be multiplied according to further reference to (*a*) scaling or (*b*) subjectivity or objectivity of data. Scaling possibilities extend from the use of covariances instead of correlation to the use of doubly standardized score matrices, but these differences have no *major* influence on the results. Subjectivity of data, i.e., the impossibility of reliability coefficients among an extended series of different observers, operates, however, to remove Q_s, O_s, etc., results from negotiability in the universe of behavioristic psychology.

If two varieties of scaling and two of data are considered along with the major variants, this creates 24 distinct factor-analytic realms of factors, which need to be distinguished by a conventional symbol system since they are never precisely equivalent and the failure to distinguish them—which is now prevalent—causes considerable confusion.

5. The choice among the three basic (P, R, T) techniques is best directed by the realm of phenomena to be investigated, while choice between the alternative transposes in each is determined by convenience, e.g., population and time available; by purpose, e.g., interest in a large or small range of variables; and by reliability of results, e.g., as to rotation, preservability, and interpretability. As to these last, Q technique is less satisfactory than R technique.

6. Q technique has its greatest usefulness in detecting and defining species types in a definitely nonhomogeneous population.[13] Where the population is reasonably homogeneous it is important to apply R technique first in order that the variables used in subsequently defining types shall be properly sampled from the total range of description, i.e., shall be factor measurements. So-called Q sort commonly gives erro-

---

[13] However, Burt argues that to determine the characteristics which distinguish types it is "more expeditious and more accurate to proceed by the analysis of variance" (5, p. 61). One should also consider here Lazarsfeld's latent subgroup analysis and Rulon's generalized discriminant function (when some basis for initial sorting exists).

neous values regarding the degree of resemblance of two individuals because equal weight is not given to the different factors in personality.

7. The primary purpose of factor analysis is to discover or confirm hypotheses as to the nature of underlying influences or dimensions. It proceeds to provide a specification equation for estimating (predicting) specific test performances, people, and occasions from the factors. It is more productive of relatively precise hypotheses than most other statistical methods and in general provides a more searching test of (deductions from) a hypothesis within a single experiment.

8. Factor analysis and analysis of variance represent respectively the multivariate and univariate approaches in scientific method. Hybrids can be formed between them, and one can with difficulty be made to perform some of the functions of the other, but essentially they are supplementary and appropriate for different phases of work and different kinds of research strategy. A retrospect on research suggests, however, that factor analysis might with advantage have been used more frequently to discover more significant measures of factors for inclusion in analysis of variance studies.

REFERENCES

1. Beebe-Center, J. *Pleasantness and unpleasantness.* New York: Century, 1933.
2. Burt, C. L. The mental differences between the sexes. *J. exp. Pedag.* 1912, **1**, 273–284.
3. Burt, C. L. *The distribution and relations of educational abilities.* London: P. S. King, 1917.
4. Burt, C. L. *Report of Consultative Committee of Board of Education on the Primary School.* Appendix III. London: H. M. Stationery Office, 1931.
5. Burt, C. L. Correlations between persons. *Brit. J. Psychol.*, 1937, **28**, 59–96.
6. Burt, C. L. A comparison of factor analysis and analysis of variance. *Brit. J. Psychol.*, Statist. Sect. 1947, **1**, 3–26.
7. Burt, C. L., *et al. How the mind works.* London: Allen and Unwin, 1933.
8. Cattell, R. B. Psychological measurement: Normative, ipsative, interactive. *Psychol. Rev.*, 1944, **51**, 292–303.
9. Cattell, R. B. *The description and measurement of personality.* New York: World Book Co., 1946.
10. Cattell, R. B. Primary personality factors in the realm of objective tests. *J. Pers.*, 1948, **16**, 459–487.
11. Cattell, R. B. The dimensions of culture patterns by factorization of national characters. *J. abnorm. soc. Psychol.*, 1949, **44**, 443–469.
12. Cattell, R. B. The main personality factors in questionnaire, self estimate material. *J. soc. Psychol.*, 1950, **31**, 3–38.
13. Cattell, R. B. *Personality: A systematic theoretical and factual study.* New York: McGraw Hill, 1950.
14. Cattell, R. B. The discovery of ergic structure in man in terms of common attitudes. *J. abnorm. soc. Psychol.*, 1950, **45**, 598–618.
15. Cattell, R. B. On the disuse and mis-

use of *P*, *Q*, *Qs*, *O*-techniques in clinical psychology. *J. clin. Psychol.*, 1951, **7**, 203–214.

16. CATTELL, R. B. *Factor analysis: An introduction and manual for the psychologist and social scientist.* New York: Harper, 1952.
17. CATTELL, R. B. The psychological dimensions of social progress in Britain, determined by *P*-technique, 1835–1935. *Hum. Relat.*, in press.
18. CATTELL, R. B., & ADELSON, M. The dimensions of social change in the U.S.A. as determined by *P*-technique. *Soc. Forces*, 1951, **30**, 190–201.
19. CATTELL, R. B., CATTELL, A. K. S., & RHYMER, R. M. *P*-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, 1947, **12**, 267–288.
20. CATTELL, R. B., & LUBORSKY, L. *P*-technique demonstrated as a new clinical method for determining personality and symptom structure. *J. gen. Psychol.*, 1950, **42**, 3–24.
21. CATTELL, R. B., & SAUNDERS, D. R. Inter-relation and matching of personality factors from behavior rating, questionnaire and objective test data. *J. soc. Psychol.*, 1950, **31**, 243–260.
22. CROSS, PATRICIA. Comparison of *P*- and *R*-technique determinations of ergic structure in common attitudes. Unpublished M.A. thesis, Univer. of Illinois, 1951.
23. HOFSTAETTER, P. R. A factorial study of cultural patterns in the U. S. *J. Psychol.*, 1951, **32**, 99–113.
24. HOLZINGER, K. J., & HARMAN, H. H. *Factor analysis.* Chicago: Univer. of Chicago Press. 1941.
25. MILLER, A. W. Confirmation of ergic structure in man, by *R*-technique analysis of objectively measured dynamic traits. Unpublished M.A. thesis, Univer. of Illinois, 1951.
26. ROGERS, C. R. *Client-centered therapy.* New York: Houghton Mifflin, 1951.
27. SAUNDERS, D. R. Practical methods in the direct factor analysis of psychological score matrices. Unpublished Ph.D. dissertation, Univer. of Illinois, 1950.
28. SPENCE, K. W. The postulates and methods of behaviorism. *Psychol. Rev.*, 1948, **55**, 67–69.
29. STEPHENSON, W. Correlating persons instead of tests. *Charact. & Pers.*, 1935, **4**, 17–24.
30. STEPHENSON, W. The inverted factor technique. *Brit. J. Psychol.*, 1936, **26**, 344–361.
31. STEPHENSON, W. The foundations of psychometry: Four factor systems. *Psychometrika*, 1936, **1**, 195–209.
32. STERN, W. *Differentielle psychologie.* Leipzig. Barth, 1911.
33. THOMSON, G. H. *The factorial analysis of human ability.* London: Univer. of London Press, 1948.
34. THURSTONE, L. L. Second order factors. *Psychometrika*, 1944, **9**, 71–100.
35. THURSTONE, L. L. The effects of selection in factor analysis. *Psychometrika*, 1945, **10**, 165–198.
36. THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.
37. WENIG, P. The relative roles of various ego defense mechanism operations in tests of misperception. Unpublished M.A. thesis, Univer. of Illinois, 1951.
38. WILLIAMS, H. V. M. A *P*-technique study of personality factors in the psychosomatic areas. Unpublished Ph.D. dissertation, Univer. of Illinois, 1949.

Received March 10, 1952.

SOME GRECO-LATIN ANALYSIS OF VARIANCE DESIGNS FOR LEARNING STUDIES[1,2]

E. JAMES ARCHER
Northwestern University[3]

In verbal learning studies having no more than about six conditions, it is frequently desirable to have the same *S* serve under all conditions. This procedure has been commonly described as having each *S* serve as his own control. To avoid a systematic bias because of a particular order, the conditions are usually arranged in a latin square. Of course, the use of such a design is defensible only when the interaction of the literal variable and rows, or literal variable and columns, will not be significant. However, this absence of interaction will occur with greater than rare frequency (see McNemar, **8**). Briefly, in a latin square: (*a*) there are as many columns and rows as there are conditions, (*b*) a particular condition appears once and only once in each row and each column, and (*c*) each row of conditions has a different permutation of orders. The analysis of variance for the latin square design has been described by Grant (**4**). Since each *S* serves in all conditions, the rows usually correspond to different *S*s. Therefore, in a nonreplicated design there are as many *S*s as there are rows. To increase the sample size there are two alternative procedures for replicating, depending upon the hypotheses under examination. One may simply repeat the same square and have a test for order of conditions or, if order is of little interest, one may use permutations of the original latin square. A design such as the latter reduces the probability of a large systematic bias due to a particular latin square. These problems of replication have been discussed by Edwards (**2, 3**).

There is still another factor to consider. In most learning studies, *S* does not learn the same material time after time under all conditions. Rather, a different task, e.g., a different list of words, is learned under each condition. These lists may be learned by all *S*s in the same order,

---

[1] The writer wishes to express his appreciation to Professor David A. Grant for his reading and very helpful criticisms of this paper.

[2] The writer is indebted to Dr. Benton J. Underwood for his helpful suggestions in the writing of this paper. The analyses herein described were originally developed for the treatment of verbal learning data obtained under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research. This project is under the direction of Dr. Underwood. The writer is also grateful to Dr. John W. Cotton for his critical reading of this paper.

[3] Now at the University of Wisconsin.

in which case list differences would be confounded with a practice effect yet to be described, or these lists also may be presented according to another "latin" square independent of the square for conditions. Such a design is commonly called a greco-latin square. In the above example Latin letters would correspond to conditions and Greek letters would correspond to lists. The analysis of variance for the greco-latin design also has been described by Grant (4).

There are two kinds of practice effects, both of which are testable with an analysis of variance of repeated measurements. The first of these is the increase in performance made by S on successive trials when learning a *single* task. This we might designate as a specific effect of practice—specific for the single task being learned. The analysis of variance of repeated measurements for this specific practice effect has been dealt with by others (1, 6, 7). The second type of practice effect is observed as a change in S's performance in the learning of several successive tasks. Here the effect of practice on several tasks belonging to the same class, e.g., lists of words, is more general than that which appears in the learning of a single task. This second case we might designate then as a general effect of practice—general for the class of tasks being learned.

It is this latter, or general, effect of practice which is testable in the proposed analyses. Since a different task is learned under a different condition, as outlined above, a systematic change in performance as a function of the ordinal sequence of combinations of conditions and tasks will be due to a general effect of practice. If, for example, the dependent variable were the number of trials required to learn the task to a given criterion, this general effect of practice would be designated a learning-how-to-learn effect.

To this point we have considered only a four-classification design: rows (Ss), columns (order of tasks), Latin letters (conditions), and Greek letters (learning materials). It is the purpose of this paper to describe analyses of variance of repeated measurements for higher orders of classification.[4]

---

[4] In a recent article (9) the suggestion was made that an analysis of variance of repeated measurements on the same subject serving under different conditions was inappropriate. The evidence presented was for a special case, however. The Ss manipulated three variations of a Müller-Lyer illusion device and were not given knowledge of their results. Since there was considerable differential transfer from one condition to the other, it is true that a repeated measurements design was inappropriate. The acquisition of a persistent perceptual bias leading to differential transfer is unlikely in the usual verbal learning experiment.

A FIVE-CLASSIFICATION ANALYSIS WITH GRECO-LATIN SQUARES ACCOUNTING FOR TWO OF THE CLASSIFICATIONS

Description of a Hypothetical Experiment

In this design repeated measurements, e.g., "days" or "trials," account for one order of classification and define this method as a repeated measurements analysis of variance. A second classification is orthogonal to that of repeated measurements. Examples of this latter classification may be an ordered variable such as hours of food deprivation, or an unordered variable such as different types of rest interval activities in a distributed practice learning problem. This second classification contains a subclass, rows, which corresponds to different *S*s. The data within this two-classification analysis may be further classified according to one or more greco-latin squares.

For these greco-latin squares, the Latin letters would account for a variable such as length of rest interval, whereas the Greek letters would account for different lists of words to be learned. Depending upon the nature of the problem, we may replicate the same greco-latin square, and consequently analyze a variance resulting from order, or we may permute the original greco-latin and reduce the probability of a systematically biased combination of Latins, Greeks, and orders. Generally we are concerned with the differences among conditions and not their order of presentation. Therefore, the discussion which follows will concern independently drawn or permuted greco-latin squares rather than a single greco-latin replicated. The reader may alter the analysis for a single greco-latin replicated following Edwards (**2**) if he wishes to test for differences among orders.

In Table 1 the schema is presented for this particular design. Six independent 4×4 greco-latin squares are presented. The numbers were taken from a table of random numbers (**5**), skipping zero entries.[5]

In this hypothetical experiment the numbers could indicate number of trials to learn a list of words to some criterion. The columns correspond to the successive days each *S* served in the experiment. The rows contain the four scores for each *S*. The blocks of three squares correspond to a particular kind of rest interval activity. The Latin letters indicate different lengths of rest intervals. The Greek letters indicate different lists of words to be learned.

[5] Although this analysis has been used successfully on real data, a hypothetical design is presented here for brevity. Some verbal learning studies under the direction of Dr. B. J. Underwood involve as many as 108 *S*s classified in three sets of 12 independently drawn 3×3 greco-latins. Presenting the original data would only unnecessarily increase the bulk of this discussion.

TABLE 1

RANDOM NUMBERS PRESENTED IN THE GENERAL FORM OF A FIVE-CLASSIFICATION* ANALYSIS WITH PERMUTED GRECO-LATIN SQUARES ACCOUNTING FOR TWO CLASSIFICATIONS

Block	*Group*	*Greco-Latin Square*				*Ss*	*Days*				*Sum*
							I	*II*	*III*	*IV*	
X	(a)	B_δ	D_β	A_γ	C_α	1	2	3	1	5	11
		A_α	C_γ	B_β	D_δ	2	7	5	4	8	24
		D_γ	B_α	C_δ	A_β	3	5	9	1	8	23
		C_β	A_δ	D_α	B_γ	4	3	7	2	5	17
						Sum	17	24	8	26	75
	(b)	A_β	D_δ	C_α	B_γ	5	9	9	3	7	28
		C_δ	B_β	A_γ	D_α	6	6	2	4	9	21
		B_α	C_γ	D_β	A_δ	7	7	8	8	6	29
		D_γ	A_α	B_δ	C_β	8	9	5	2	3	19
						Sum	31	24	17	25	97
	(c)	D_β	B_α	A_δ	C_γ	9	3	6	7	4	20
		A_γ	C_δ	D_α	B_β	10	4	5	5	4	18
		B_δ	D_γ	C_β	A_α	11	5	5	5	4	19
		C_α	A_β	B_γ	D_δ	12	3	1	5	3	12
						Sum	15	17	22	15	69
Y	(a)	A_β	B_δ	D_α	C_γ	13	7	4	3	5	19
		B_γ	A_α	C_δ	D_β	14	8	9	6	1	24
		C_α	D_γ	B_β	A_δ	15	1	1	8	3	13
		D_δ	C_β	A_γ	B_α	16	7	4	4	1	16
						Sum	23	18	21	10	72
	(b)	A_δ	D_α	B_γ	C_β	17	9	6	2	2	19
		D_β	A_γ	C_α	B_δ	18	1	3	4	3	11
		C_γ	B_β	D_δ	A_α	19	1	4	8	7	20
		B_α	C_δ	A_β	D_γ	20	1	6	3	5	15
						Sum	12	19	17	17	65
	(c)	A_α	B_δ	C_β	D_γ	21	3	2	4	4	13
		D_δ	C_α	B_γ	A_β	22	3	6	2	2	13
		B_β	A_γ	D_α	C_δ	23	3	5	5	1	14
		C_γ	D_β	A_δ	B_α	24	3	2	2	1	8
						Sum	12	15	13	8	48

In this particular design each *S* serves in only one row. Consequently, 24 *S*s have participated in this hypothetical experiment. In addition to isolating the sums of squares of each of these main effects, we may expect three simple interactions: (*a*) Columns × Blocks, (*b*) Latins

---

* For the purposes of the two five-classification analyses the sixth classification of Group will be ignored. This classification will be used when the data are analyzed according to a six-classification procedure.

×Blocks, and (*c*) Greeks×Blocks. Ordinarily in a latin or greco-latin square design we would not expect to have an interaction for the Latin or Greek letters with a main effect, but in this design we have an amalgamation of the usual greco-latin design into a factorial design; hence the interactions.

General Form of Analysis

The general equation for this analysis is as follows:

$$\begin{aligned}\sum (X_i - \overline{\overline{X}})^2 = {} & n_C n_{S/B} \sum (\overline{X}_B - \overline{\overline{X}})^2 + n_S \sum (\overline{X}_C - \overline{\overline{X}})^2 \\ & + n_S \sum (\overline{X}_L - \overline{\overline{X}})^2 + n_S \sum (\overline{X}_\Gamma - \overline{\overline{X}})^2 \\ & + \sum^{B} n_C \sum^{S} (\overline{X}_S - \overline{X})^2 + n_{S/B} \sum (\overline{X}_{BC} - \overline{X}_B \\ & - \overline{X}_C + \overline{\overline{X}})^2 + n_{S/B} \sum (\overline{X}_{BL} - \overline{X}_B - \overline{X}_L + \overline{\overline{X}})^2 \qquad [1] \\ & + n_{S/B} \sum (\overline{X}_{B\Gamma} - \overline{X}_B - \overline{X}_\Gamma + \overline{\overline{X}})^2 \\ & + \sum^{B} \sum^{S} (X_i - \overline{X}_C - \overline{X}_L - \overline{X}_\Gamma - \overline{X}_S + 3\overline{X}_B)^2\end{aligned}$$

where X_i = any single score

$\overline{\overline{X}}$ = Grand mean

$\overline{X}_B$ = Block mean

$\overline{X}_C$ = Column mean

$\overline{X}_L$ = Latin mean

$\overline{X}_S$ = Subject mean

$\overline{X}_\Gamma$ = Greek mean

$\overline{X}_{BC}$ = cell mean based on Block-Column classification

$\overline{X}_{BL}$ = cell mean based on Block-Latin classification

$\overline{X}_{B\Gamma}$ = cell mean based on Block-Greek classification

$\tilde{n}_C$ = number of Columns

$\tilde{n}_{S/B}$ = number of *S*s within block

n_S = total number of *S*s

n_B = number of Blocks

N = total number of scores

It is to be understood that when a summation sign with a limiting superscript precedes another summation sign with a limiting superscript, the first superscript governs the limit of the second and all others within a given quantity, e.g.,

$$\sum^{G} \left[\sum^{S} (\sum X)^2 / n_C \right] \quad \text{means} \quad \sum^{G} \left[\sum^{S/G} (\sum X) / n_C \right].$$

The degrees of freedom for the general case would be analyzed as in Table 2. The *df* for the Pooled Ss×Columns is shown as $n_B(n_{S/B}n_C - n_{S/B} - n_C - n_L - n_\Gamma + 3)$. This appears as if the value within the parentheses is multiplied by the number of blocks. Although this procedure is admissible when there are an equal number of Ss within all blocks, it is to be understood that the *df* associated with these Ss×Columns is actually obtained by summating for all blocks. This is of theoretical significance only with respect to *df* but of practical significance with respect to the corresponding sums of squares.

TABLE 2

DISTRIBUTION OF DEGREES OF FREEDOM FOR BASIC FIVE-CLASSIFICATION DESIGN

Source of Variation	*df*
Blocks	$n_B - 1$
Ss/Blocks	$n_B(n_{S/B} - 1)$
Columns	$n_C - 1$
Columns×Blocks	$(n_C - 1)(n_B - 1)$
Latins	$n_L - 1$
Latins×Blocks	$(n_L - 1)(n_B - 1)$
Greeks	$n_\Gamma - 1$
Greeks×Blocks	$(n_\Gamma - 1)(n_B - 1)$
Pooled Ss×Columns/Blocks	$n_B(n_{S/B}n_C - n_{S/B} - n_C - n_L - n_\Gamma + 3)$
Total	$(n_{S/B})(n_B)(n_C) - 1$

The computational formulae are as follows:

(1) Correction factor $= C = (\sum X)^2/N$

(2) Total $SS = \sum X^2 - C$

(3) Blocks $SS = \sum^{B} (\sum X)^2/(n_{S/B})(n_C) - C$

(4) Ss/Blocks $SS = \sum^{B}\left[\sum^{S} (\sum X)^2/n_C - (\sum X)^2/(n_{S/B})(n_C)\right]$

(5) Columns $SS = \sum^{C} (\sum X)^2/n_S - C$

(6) Latins $SS = \sum^{L} (\sum X)^2/n_S - C$

(7) Greeks $SS = \sum^{\Gamma} (\sum X)^2/n_S - C$

(8) Columns × Blocks $SS = \sum^{CB} (\sum X)^2/n_{S/B} - \sum^{C} (\sum X)^2/n_S$
$$- \sum^{B} (\sum X)^2/(n_{S/B})(n_C) + C$$

(9) Latins × Blocks $SS = \sum^{LB} (\sum X)^2/n_{S/B} - \sum^{L} (\sum X)^2/n_S$

$$- \sum^{B} (\sum X)^2/(n_{S/B})(n_C) + C$$

(10) Greeks × Blocks $SS = \sum^{\Gamma B} (\sum X)^2/n_{S/B} - \sum^{\Gamma} (\sum X)^2/n_S$

$$- \sum^{B} (\sum X)^2/(n_{S/B})(n_C) + C$$

(11) Pooled *S*s × Columns/Blocks *SS*

$$= \sum^{B} \Big[\sum X^2 - \sum^{C} (\sum X)^2/n_{S/B} - \sum^{L} (\sum X)^2/n_{S/B}$$
$$- \sum^{\Gamma} (\sum X)^2/n_{S/B} - \sum^{S} (\sum X)^2/n_C + 3(\sum X)^2/(n_{S/B})(n_C) \Big]$$

Analysis of Variance of Hypothetical Data: Basic Five-Classification Design

Using the computational formulae given in the preceding section the total variance of the data from the hypothetical experiment shown in

TABLE 3

ANALYSIS OF VARIANCE OF DATA FROM HYPOTHETICAL EXPERIMENT: BASIC FIVE-CLASSIFICATION DESIGN

Source of Variation	*Sum of Squares*	*df*	*Mean Square*	*F*
Independent Observations				
Blocks	32.667	1	32.667	5.462*
Pooled *S*s/Blocks	136.458	22	6.203	
Total between subjects	169.125	23		
Correlated Observations				
Columns	9.375	3	3.125	—
Columns×Blocks	25.750	3	8.583	1.435
Latins	20.875	3	6.958	1.240
Latins×Blocks	8.583	3	2.861	—
Greeks	10.875	3	3.625	—
Greeks×Blocks	3.916	3	1.305	—
Pooled *S*s×Columns/Blocks	303.126	54	5.613	
Total within subjects	382.500	72		
Total for experiment	551.625	95		

* For *df* 1 and 22, $p_{.05}F=4.30$; $p_{.01}F=7.94$.

Table 1 was analyzed. This analysis is shown in Table 3 in the format suggested by Edwards (3). The advantage of this arrangement of component values is that the appropriate error terms become immediately obvious.

In terms of the hypothetical variables outlined earlier, we would conclude that there was a significant difference between the two rest interval activities as indicated by an *F* of 5.462 for Blocks.[6] The Pooled *S*s/Blocks[7] is the proper error term for testing the Block means' estimate of population variance. This is true only if the *S*s were randomly assigned to the Blocks, however. Such an assumption is made here.

A test for individual differences could be made by dividing the Pooled *S*s/Blocks variance estimate by the Pooled *S*s×Columns/Blocks.[8]

The proper error term for the remaining variance estimates, based as they are upon correlated observations, will be the Pooled Rows ×Columns/Blocks. According to our hypothetical assignment of variables to coordinates, we find that none of the main effects is significant. Also none of the interactions of these main effects and rest-interval activities, Blocks, is significant.

Limitations

As may be seen through the use of an experimental design such as outlined above, a great deal of information can be obtained even though only 24 *S*s were involved. In this example it was possible to evaluate the effect of four main variables and three simple interactions. Such a return strongly supports Grant's prediction (4, p. 441) that the [greco]-latin square design will find wide application in psychology. Nevertheless this design has certain limitations which should be considered. Although each datum is governed by four classifications, it does not yield interactions between all main effects. This could be a serious drawback if, for example, we wanted to test the significance of the length of rest interval × practice interaction. According to the above proposed design such an interaction, Columns × Latins, would not be isolable, and, in fact, is assumed to be not significant. However, the difficulty could be partially circumvented by the judicious assignment of variables to coordinates. Before the experiment was performed *E* could decide which interaction tests would be important and assign his variables accordingly. For example, if we wish to test the significance of a practice×length of rest interval, we could assign "length of rest interval" to the coordinate occupied by "rest interval activities." We would lose the test for a rest interval activity×practice interaction, but we would still have a test for the main effect.[9]

---

[6] Theoretically an *F* ratio of this magnitude should occur but once in about 20 drawings of samples of random numbers of this size.

[7] Logically, of course, the Pooled *S*s/Blocks can be used as the error term only if homogeneity of variance is demonstrable. When there are only two cases involved a two-tail *F* test will be shorter than a Bartlett test.

[8] Now the assumption of homogeneity of variance must be met for the Pooled *S*s×Columns/Blocks. If the numerator does not satisfy such an assumption, each contributing Block would be tested separately.

A FIVE-CLASSIFICATION ANALYSIS OF VARIANCE WITH GRECO-LATIN SQUARES ACCOUNTING FOR TWO CLASSIFICATIONS AND THE LEARNING TASKS DEFINING A COORDINATE

This design is essentially a variation of the one discussed previously, but because of its wide applicability it deserves individual consideration. In this design the "nature" of the learning task is dimensionalized, i.e., the learning tasks define points along one of the coordinates. For example, with reference to Table 1, the Columns (days), Rows (*S*s), and Latins (lengths of rest intervals) remain the same but the Blocks represent different degrees of intratask similarity. Let us assume Block *X* has high intratask similarity and Block *Y* low intratask similarity. What this really means is that the four lists which are learned by *S*s in Block *X* will be so constructed as to have high intralist similarity and the four lists which are learned by *S*s in Block *Y* will have low intralist similarity. In brief the "nature" of the learning tasks (Greeks) *defines* the main coordinate (Blocks).

This variation introduces a few modifications into the analysis of the total variance, the associated *df*, and the choice of error terms for *F* tests. Since different learning materials are learned by the *S*s in each block, a significant difference between blocks would mean that the materials were truly different, or more precisely, that the "nature" of differences between blocks influenced the measure of learning. Such a finding could be of considerable importance for some hypotheses. However, a significant difference between material classified by Blocks would not be equivalent to a difference between materials classified by Greeks. In fact, following the first analysis, a difference between Greeks would be confounded by a difference between Blocks. The solution to this dilemma is to test for differences among Greeks within each Block. There will be as many *F* tests for Greeks as there are Blocks. A significant *F* ratio would mean the materials within (and defining) a Block were different with respect to the dependent variable measured. This variation in analysis alters equation (1) by eliminating a Greek × Block interaction and accounts for Greek variation in terms of deviation from Block mean rather than Grand mean. The altered equation is as follows:

---

[9] This manipulation would give us four blocks and the number of rest interval activities would have to be increased to be classified as a Latin or Greek.

$$\begin{aligned}\sum (X_i - \bar{\bar{X}})^2 = {} & n_C n_{S/B} \sum (\bar{X}_B - \bar{\bar{X}})^2 + n_S \sum (\bar{X}_C - \bar{\bar{X}})^2 \\ & + n_S \sum (\bar{X}_L - \bar{\bar{X}})^2 + \sum^{B} n_S \sum (\bar{X}_\Gamma - \bar{X}_B)^2 \\ & + \sum^{B} n_C \sum^{S} (\bar{X}_S - \bar{X}_B)^2 + n_{S/B} \sum (\bar{X}_{CB} - \bar{X}_C \\ & - \bar{X}_B + \bar{\bar{X}})^2 + n_{S/B} \sum (\bar{X}_{LB} - \bar{X}_L - \bar{X}_B + \bar{\bar{X}})^2 \\ & + \sum^{B} \sum^{S} (X_i - \bar{X}_C - \bar{X}_L - \bar{X} \;\; - \bar{X}_S + 3\bar{X}_B)^2 .\end{aligned} \qquad [2]$$

The breakdown for *df* is also slightly altered and is shown in Table 4.

TABLE 4

DISTRIBUTION OF DEGREES OF FREEDOM FOR VARIATION OF FIVE-CLASSIFICATION DESIGN

Source of Variation	*df*
Blocks	$n_B - 1$
Pooled *S*s/Blocks	$n_B(n_{S/B} - 1)$
Columns	$n_C - 1$
Columns × Blocks	$(n_C - 1)(n_B - 1)$
Latins	$n_L - 1$
Latins × Blocks	$(n_L - 1)(n_B - 1)$
Pooled Greeks/Blocks	$n_B(n_\Gamma - 1)$
Pooled *S*s × Columns/Blocks	$n_B(n_{S/B} n_C - n_{S/B} - n_C - n_L - n_\Gamma + 3)$
Total	$(n_{S/B})(n_C)(n_B) - 1$

The computational formulae remain the same as for the previous analysis except that:

Pooled Greeks

$$SS = \sum^{B} \left[\sum^{\Gamma} (\sum X)^2 / n_{S/B} - (\sum X)^2 / n_C(n_{S/B}) \right]$$

replaces formulae (7) and (10).

Analysis of Hypothetical Data: A Variation with Greeks Defining a Coordinate

The significant change in re-analyzing the data presented in Table 1 is that the Greek × Block interaction is absent and more F tests are in order. The analysis of variance is presented in Table 5.[10] All sources of variance are tested and interpreted as before except the Greeks/Blocks, each of which is tested for significance by dividing its variance estimate by its corresponding $Ss \times$Column/Block, e.g., Greeks/Block$_X$ divided by $Ss \times$Columns/Block$_X = 0.443$. In both cases in Table 5 the ratio was less than 1.000. This would be interpreted to mean, in our hypothetical experiment, that the four lists in each of the two blocks did not differ significantly one from the other in learning difficulty. That the two *sets* of four lists were of unequal difficulty is indicated by the F ratio of 5.462 for the Blocks divided by Pooled Ss/Blocks.

TABLE 5

Analysis of Variance of Data from Hypothetical Experiment: A Variation in which Learning Task Defines a Coordinate

Source of Variation	*Sum of Squares*	*df*	*Mean Square*	*F*
Independent Observations				
Blocks	32.667	1	32.667	5.462*
Pooled *Ss*/Blocks	136.458	22	6.203	
Total between subjects	169.125	23		
Correlated Observations				
Columns	9.385	3	3.125	—
Columns × Blocks	25.750	3	8.583	1.435
Latins	20.875	3	6.958	1.240
Latins × Blocks	8.583	3	2.861	—
Pooled Greeks/Blocks	14.791	6		
Greeks/Block *X*	6.562	3	2.187	—
Greeks/Block *Y*	8.229	3	2.743	—
Pooled *Ss* × Columns/Blocks	303.126	54	5.613	
Ss × Columns/Block *X*	133.230	27	4.934	
Ss × Columns/Block *Y*	169.896	27	6.292	
Total within subjects	382.500	72		
Total for experiment	551.625	95		

* For *df* 1 and 22, $p_{.05}F = 4.30$; $p_{.01}F = 7.94$.

---

[10] This analysis has also been applied to real data but for convenience and continuity of comparison the same hypothetical data will be used again.

A SIX-CLASSIFICATION ANALYSIS OF VARIANCE WITH GRECO-LATIN SQUARES ACCOUNTING FOR TWO CLASSIFICATIONS

This analysis is like the first except that another coordinate has been added. Such an addition considerably increases the complexity of the analysis and in one respect also reduces its precision somewhat. The same hypothetical data given in Table 1 will be used in this example. Now, however, the classification of Group will be considered. These Groups are designated by the letters *a*, *b*, and *c* in Table 1. The hierarchy is: Blocks contain Groups, Groups contain Squares, Squares contain Rows, and Rows contain scores, each of which lies in a Column. Finally the conditions under which each score is obtained are governed by a pair of letters, one Latin and one Greek.

Although this added coordinate increases our information from the data, we sacrifice some precision in that one variance estimate is based upon a pool of three triple interactions. At present these triple interactions are not isolable. The Pooled Triple Interactions are determined by subtraction. This is not too serious a disadvantage since if one of the three interactions is significant and if the other two are not significantly restricted, the Pooled Triple Interaction divided by the Pooled

TABLE 6

DISTRIBUTION OF DEGREES OF FREEDOM FOR BASIC SIX-CLASSIFICATION DESIGN

Source of Variation	*df*
Blocks	n_B-1
Groups	n_G-1
Blocks × Groups	$(n_B-1)(n_G-1)$
Pooled *S*s/Groups/Blocks	$(n_{S/G}-1)(n_G)(n_B)$
Columns	n_C-1
Columns × Blocks	$(n_C-1)(n_B-1)$
Columns × Groups	$(n_C-1)(n_G-1)$
Latins	n_L-1
Latins × Blocks	$(n_L-1)(n_B-1)$
Latins × Groups	$(n_L-1)(n_G-1)$
Greeks	$n_\Gamma-1$
Greeks × Blocks	$(n_\Gamma-1)(n_B-1)$
Greeks × Groups	$(n_\Gamma-1)(n_G-1)$
Columns × Blocks × Groups	$(n_C-1)(n_B-1)(n_G-1)$
Latins × Blocks × Groups	$(n_L-1)(n_B-1)(n_G-1)$
Greeks × Blocks × Groups	$(n_\Gamma-1)(n_B-1)(n_G-1)$
Pooled *S*s × Columns/Groups/Blocks	$n_Bn_G(n_{S/G}n_C-n_{S/G}-n_C-n_L-n_\Gamma+3)$
Total	$(n_{S/G})(n_C)(n_G)(n_B)-1$

Ss×Columns/Groups will indicate this significance. Deciding which of the triple interactions is significant is even less precise. For the present a rational analysis must be used. Although these disadvantages appear almost to invalidate this design, it should be pointed out that a significant and important triple interaction is infrequent. Furthermore, in almost all cases the triple interaction involving the learning materials (Greeks) will be insignificant and unimportant since an effort usually will be made to attain uniformity in tasks.

General Form of Analysis

The distribution of the *df* for this design is shown in Table 6. Again, as in the five-classification analysis, it is to be noted that the Pooled *Ss*×Columns/Groups/Blocks is actually obtained by summing these interactions for Groups and Blocks and not simply by multiplying by $n_B n_G$ as indicated. For the *df* this is of theoretical significance, but for the sums of squares this concept is of practical importance.

The computational formulae for this six-classification analysis is as follows:

$$(1)\ \text{Correction factor} = C = (\sum X)^2/N$$

$$(2)\ \text{Blocks } SS = \sum^{B} (\sum X)^2/(n_{S/B})(n_C) - C$$

$$(3)\ \text{Groups } SS = \sum^{G} (\sum X)^2/n_{S/G})(n_C)(n_B) - C$$

$$\begin{aligned}(4)\ \text{Blocks} \times \text{Groups } SS = &\sum^{BG} (\sum X)^2/n_{S/G})(n_C)\\ &- \sum^{B} (\sum X)^2/(n_{S/B})(n_C)\\ &- \sum^{G} (\sum X)^2/(n_{S/G})(n_C)(n_B) +\end{aligned}$$

$$\begin{aligned}(5)\ &Ss/\text{Groups/Blocks } SS\\ &= \sum^{B} \sum^{G} \left[\sum^{S} (\sum X)^2/(n_C) - (\sum X)^2/(n_{S/G})(n_C) \right]\end{aligned}$$

$$(6)\ \text{Columns } SS = \sum^{C} (\sum X)^2/n_S - C$$

$$\begin{aligned}(7)\ \text{Columns} \times \text{Blocks } SS = &\sum^{CB} (\sum X)^2/n_{S/B} - \sum^{C} (\sum X)^2/n_S\\ &- \sum^{B} (\sum X)^2/(n_{S/B})(n_C) + C\end{aligned}$$

(8) Columns × Groups $SS = \sum^{CG} (\sum X)^2/(n_{S/G})(n_B) - \sum^{C} (\sum X)^2/n_S$

$$- \sum^{G} (\sum X)^2/n_{S/G})(n_C)(n_B) + C$$

(9) Latins $SS = \sum^{L} (\sum X)^2/n_S - C$

(10) Latins × Blocks $SS = \sum^{LB} (\sum X)^2/n_{S/B} - \sum^{L} (\sum X)^2/n_S$

$$- \sum^{B} (\sum X)^2/(n_{S/B})(n_C) + C$$

(11) Latins × Groups $SS = \sum^{LG} (\sum X)^2/(n_{S/G})(n_B) - \sum^{L} (\sum X)^2/n_S$

$$- \sum^{G} (\sum X)^2/(n_{S/G})(n_C)(n_B) + C$$

(12) Greeks $SS = \sum^{\Gamma} (\sum X)^2/n_S - C$

(13) Greeks × Blocks $SS = \sum^{\Gamma B} (\sum X)^2/n_{S/B} - \sum^{\Gamma} (\sum X)^2/n_S$

$$- \sum^{B} (\sum X)^2/n_{S/B})(n_C) + C$$

(14) Greeks × Groups $SS = \sum^{\Gamma G} (\sum X)^2/(n_{S/G})(n_B) - \sum^{\Gamma} (\sum X)^2/n_S$

$$- \sum^{G} (\sum X)^2/(n_{S/G})(n_C)(n_B) + C$$

(15) Pooled *S*s × Columns/Groups/Blocks SS

$$= \sum^{B} \sum^{G} \Big[\sum X^2 - \sum^{S} (\sum X)^2/n_{C/G} - \sum^{C} (\sum X)^2/n_{S/G}$$

$$- \sum^{L} (\sum X)^2/n_{S/G} - \sum^{\Gamma} (\sum X)^2/n_{S/G} + 3(\sum X)^2/(n_{S/G})(n_C) \Big]$$

(16) Total $SS = \sum X^2 - C$

(17) Pooled Triple Interaction

$$SS = (16) - (2) + (3) + (4) + \cdots + (15)$$

Although this is a six-classification design there are no interactions beyond three triple interactions. If this Pooled Triple Interaction does not prove to be significantly greater than the Pooled *S*s×Columns /Groups/Blocks, these two *SS* may be combined and divided by the combined *df*'s and used as a more reliable estimate of error variance. Throughout these analyses it is assumed that before two or more variance estimates are "pooled," the proposed pool meets a test for homogeneity of variance.

Analysis of Variance of Hypothetical Data: Basic Six-Classification Design

If as in Table 1 we let Blocks correspond to two different kinds of rest-interval activities, Groups correspond to three levels of work rate for these activities and Columns, Latins, and Greeks correspond to days, lengths of rest intervals, and learning materials, respectively, we are ready to re-analyze the data according to the proposed six-classification analysis. This analysis is presented in Table 7. Since the Pooled

TABLE 7

ANALYSIS OF VARIANCE OF DATA FROM HYPOTHETICAL EXPERIMENT: BASIC SIX-CLASSIFICATION DESIGN

Source of Variation	*Sum of Squares*	*df*	*Mean Square*	*F*
Independent Observations				
Blocks	32.667	1	32.667	6.515*
Groups	32.813	2	16.407	3.272
Blocks × Groups	13.395	2	6.698	1.336
Pooled *Ss*/Groups/Rows	90.248	18	5.014	
Total between subjects	169.123	23		
Correlated Observations				
Columns	9.375	3	3.125	—
Columns × Blocks	25.750	3	8.583	1.499
Columns × Groups	20.687	6	3.448	—
Latins	20.875	3	6.958	1.216
Latins × Blocks	8.583	3	2.861	—
Latins × Groups	60.937	6	10.156	1.774
Greeks	10.875	3	3.625	—
Greeks × Blocks	3.916	3	1.305	—
Greeks × Groups	15.437	6	2.573	—
Pooled Triple Interactions:				
Columns × Blocks × Groups		18		
Latins × Blocks × Groups	122.311			
Greeks × Blocks × Groups	206.067	36	5.724	
Pooled *Ss* × Columns/Groups/Blocks	83.756	18		
Total within subjects	382.500	72		
Total for experiment	551.625	95		

* For df 1 and 22, $p_{.05}F=4.30$; $p_{.01}F=7.94$.

Triple Interaction was not significantly greater than the Pooled Ss ×Columns/Groups/Blocks ($F=1.460$, 2.22 required for $p_{.05}$) these two variance estimates have been combined and used as an "error" term for testing the Columns, Latins, and Greeks and their interactions. As in the previous analyses, only the Block means are significantly different.

The interpretations and rationale of the tests remain the same as before. It is understood of course that the use of one greco-latin square per Group is only the minimum. Any multiple number of squares may be used as required by the reliability desired and other features of the experimental design. If one wished, he could derive a variation of this design with Greeks defining Blocks as previously described.

Limitations and Applications

The chief limitation of this design is the loss of precision in isolating the sums of squares for each of the triple interactions. This could be a serious deterrent in some studies. In the usual learning studies it is unlikely that the triple interactions will be of theoretical significance. In such cases a six-classification design could be very useful since much information can be gleaned from relatively little data.

Both the five- and six-classification analyses and their variations have wide applicability, assuming the nonsignificance of nonisolable interactions. In addition to those already discussed, examples of variables which could be tested by using the basic design in which the same materials are learned in all blocks are: methods of material presentation (audio, visual, audio and visual), rates of material presentation, levels of frustration induction, levels of motivation, e.g., ego-involvement studies, hours of sleep deprivation. Other variables which could be tested by using the variation design in which materials define blocks are: levels of meaningfulness, levels of affectivity, different types of presentation (paired associates versus verbal discrimination).

Summary

Two basic multiple-classification analyses of variance of repeated measurements incorporating greco-latin squares were described in detail. A useful variation of one of the designs was also described. In each case these analyses were applied to 96 random numbers which were arranged in the appropriate experimental designs under discussion. The applicability and limitations of the methods were described.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. EDWARDS, A. L. Homogeneity of variance and the Latin square design. *Psychol. Bull.*, 1950, **47**, 118–129.
3. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. GRANT, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
5. KENDALL, M. G., & SMITH, B. B. Randomness and random sampling numbers. *J. roy. statist. Soc.*, 1938, **101**, 147–166.
6. KOGAN, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.
7. LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
8. MCNEMAR, Q. On the use of latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398–401.
9. SMITH, B. B. On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments. *Brit. J. Psychol.*, 1951, **42**, 250–268.

Received January 2, 1952.

A NOTE ON PROFILE SIMILARITY

HAROLD WEBSTER
University of Kentucky

The article by Osgood and Suci (2) on a measure of profile similarity deserves comment. The need for such measures is certainly great, as the authors point out, and their treatment of the problem is interesting. This note is intended to suggest other approaches to the same problem.

It is true that most correlation coefficients, when used as measures of profile similarity, disregard the absolute differences between means of profiles. One exception, however, is intraclass r. Intraclass r between two parallel profiles decreases in value from $+1.00$ toward zero as the profiles are moved farther apart. It has the advantage of having a known standard error, but the disadvantage of possessing a restricted negative range (1).

The problem concerning the orthogonality of measurement variates should be made more explicit. It is true that if measurement variates are orthogonal, that is to say independent (and thus uncorrelated), then D has the value given by the authors. In any random sampling problem, however, it is unlikely that five measures such as those in their Fig. 1 would have zero intercorrelations. The problem of allowing for the effects of such intercorrelations has been a central one in multivariate analysis. Unfortunately the complete solution for comparing profiles is indeterminate when each profile represents only one person and hence provides only a single degree of freedom. If each profile should represent means of persons in a small group, then D could be obtained from

$$D^2 = \sum_{i}^{k} \sum_{j}^{k} a^{ij} d_i d_j \qquad (i, j = 1, \cdots, k),$$

where a^{ij} is the matrix inverse to a_{ij}, the measurement variate dispersion matrix for the entire sample, and d_i, d_j are differences between means for the two groups. This is Mahalanobis' D^2, a statistic which not only allows for the effects of the intercorrelations among the k variates but also may be tested for significance by an F ratio (3).

Rao (4) has given a transformation for the original measurement variates which makes them mutually independent. This transformation is lengthy if there are more than six or eight variates, but has the advantage of avoiding the even lengthier inversion of the dispersion matrix which is ordinarily necessary for computing the D^2. With this transformation Mahalanobis' D becomes the D used by Osgood and Suci.

REFERENCES

1. FISHER, R. A. *Statistical methods for research workers*. London: Oliver and Boyd, 1948.
2. OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, 49, 251–262.
3. RAO, C. R. Tests of significance in multivariate analysis. *Biometrika*, 1948, **35**, 58–79.
4. RAO, C. R. On a transformation useful in multivariate computations. In P. C. Mahalanobis, D. N. Majumdar, *et al.*, *Anthropometric survey of the United Provinces, 1941: A statistical survey*. *Sankhyā*, 1949, **9**, 89–324. Appendix 5, pp. 251–253.

Received June 21, 1952.

ON THE RELIABILITY OF THE LEADERLESS GROUP DISCUSSION TECHNIQUE

BORIS SEMEONOFF
University of Edinburgh

Ansbacher's recent survey *The History of the Leaderless Group Discussion Technique* (1) quotes only one reference to work on reliability of the technique, that of Bell and French (2). Evaluation was in terms of mutual ranking "in order of preference for discussion leader." Although the relevance of this criterion is not altogether clear, the work did at any rate yield an average correlation of mean ranks for each candidate, within the various groups in which he participated, of 0.75. This must be accepted as a positive result, although regarded as a reliability coefficient it is not outstandingly high.

The object of this note is to call attention to an alternative method of evaluation, based not on mutual rating, but on an objective assessment of participation.

The underlying principle is similar to that of the sociometric method of Moreno and others. Essentially it is based on the recognition of two types of leadership in group discussion, which may be designated, respectively, as *active* and *passive*. The former may be measured by the number of times a member of the group speaks, and the latter by the number of times he is spoken to. It would perhaps be feasible to evolve a formula combining these two variables which might give an optimum estimate of effectiveness in the discussion situation, although it will of course be recognized that effectiveness cannot be measured in quantitative terms alone. However, the derivation of such a formula is not relevant to the present purpose.

Work along these lines was carried out by the author in connection with a selection procedure similar to that described by the OSS Staff (4). No observer participated in the discussion, or entered the circle in which the candidates sat, except to explain the procedure at the outset. The discussion lasted about forty minutes, divided into two approximately equal periods, with a different topic for each period. The progress of the discussion was recorded in such a way as to show not only the number of times each member of the group spoke or was spoken to, but also the sequence of the various contributions to the discussion. This record, however, was kept purely for research purposes, and was not used in rating effectiveness in the group discussion situation; such rating was carried out according to principles common to most of the writers referred to by Ansbacher.

From the record four "raw" measures may be obtained, and regarded as two measures each of "active" and "passive" leadership in the group discussion situation, as defined above. These are referred to in what follows as AI, AII, PI, and PII, the respective topics being designated as I and II.

If high correlations are found between AI and AII and between PI and PII, we may conclude that reliability of these measures has been demonstrated, and consequently that a meaningful formula for "leadership" in this situation may also yield a reliable measure. In other words, there would appear to be a relatively stable structure within the group when engaged in this particular activity. If, on the other hand, correlations between AI and PI and between AII and PII tend to be higher than the reliability coefficients for A and P, it would suggest that any hierarchical structure within the group during the discussion of a particular topic was more closely related to such factors as knowledge (or even only knowledgeability) in the relevant field, than to any permanent structure within the group as a group.

Detailed results have not been published, and are not now available, but the second of the possibilities outlined above was that which emerged in nearly every case. Reliability was thus seen to be low, seldom rising above the level of a correlation of the order of 0.4. This finding would also appear to cast serious doubt on the validity for practical purposes of a discussion based on a single randomly chosen topic, since knowledge or interest would appear to be a dominant factor. It would also appear to be an argument in favor of what may be described as the "object-stimulus" method of group discussion latterly adopted by War Office Selection Boards in Great Britain. This method makes use of a freely developing discussion based on an informal conversational opening similar to one which might be used among a group of strangers in a railway carriage. Reference to this method is made by Harris (**3**), but it is hoped that a further discussion of its rationale will be published later. Whether such a device leads to a more reliable discussion situation has not, so far as the writer knows, been investigated.

REFERENCES

1. Ansbacher, H. L. The history of the leaderless group discussion technique. *Psychol. Bull.*, 1951, **48**, 383–391.
2. Bell, B., & French, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 764–767.
3. Harris, H. *The group approach to leadership testing.* London: Routledge & Kegan Paul, 1949.
4. OSS Assessment Staff. *Assessment of men.* New York: Rinehart, 1948.

Received March 10, 1952.

GENETICS OF SCHIZOPHRENIA: A REJOINDER

NICHOLAS PASTORE
Queens College

Since space limitations do not permit a detailed answer to Hurst's reply (1) to my review (3) of Kallmann's *Genetics of Schizophrenia* (2) I shall select a few of the major points for discussion. Most of the inaccuracies of understanding and misquotations which Hurst introduces in his reply will have to be ignored.

1. Hurst asserts that Haldane and Hogben acclaimed Kallmann's work. In response to an inquiry by the writer Haldane wrote: "I regret that I possess no copy of the monograph to which you refer, and have no recollection of having ever read it, or commented on it any way, favourably or otherwise. Mr. Hurst may, of course, attribute this blank in my mind to senile dementia, Freudian repression, or some other failing on my part. From what little I have read on the genetics of schizophrenia it appears to me that its genetical investigation presents grave difficulties. In view of its variable age of onset and uncertain diagnosis I am sceptical as to any statements made about its genetics."

2. Despite Hurst's claim to the contrary, Kallmann did include the "doubtful" schizophrenics in the "definite" schizophrenic category in many of his analyses. My statement concerning the 16.4 per cent figure (and other figures as well) as including the "doubtfuls" was based on computations in accordance with Kallmann's procedure. Since it can be arithmetically shown that the "doubtfuls" were included in Kallmann's theoretical discussions my criticism of Kallmann's demonstration of a "gene-coupling" between the hereditary dispositions of tuberculosis and schizophrenia still holds. (The relevant calculations can be made available to the interested reader.)

3. In contradiction to my statement in the review Hurst asserts that Kallmann "invariably" included numbers as well as percentages in the subgroups. Hurst is in error. Kallmann's Table 10 does not state the numbers for *any* category with the exception of the second column entitled "absolute." Kallmann's Table 38 does not state the numbers for any cell. Other examples exist.

4. Hurst alleges that I failed to understand Kallmann's Table 58 because I did not grasp the significance between "expectancy" and "net" figures. This table deals with the health outcome of those children who have parents both of whom are schizophrenic. According to Kallmann, the probability that such offspring will develop schizophrenia is 68.1 per cent and the probability that such offspring will develop schizoidia is 45.7 per cent. I criticized these probabilities because their sum is significantly larger than unity whereas the sum of independent probabilities, mathematically speaking, would not exceed unity (Kallmann

presents no computation of the probability that the offspring will remain healthy). Therefore, I inferred the presence of "overlapping" categories. This may or may not be the case. The absurd result of a sum of apparently mutually exclusive (and not even exhaustive) probabilities which exceeds unity must be explained. Hurst fails to offer an explanation.

5. In my original review I demonstrated a numerical contradiction between the 108 secondary cases enumerated in Chapter 5 of Kallmann's book and the enumeration of the same cases in Tables 34–37. Hurst alleges that I am in error because of my apparent failure to distinguish between "expectancy" and "net" figures. However, the distinction between the two types of figures is not involved and the discrepancy is a real one, as the reader can verify for himself by summing the "absolute" numbers in the relevant tables. This point is adequately discussed in the original review.

6. In criticism of Kallmann's Table 10, I stated that "overlapping categories" were involved (**3**, p. 294). Hurst states that this is not the case (**1**, p. 407). Let us examine this table. The subcolumn headings entitled "schizophrenia in one parent" and the other entitled "schizophrenia in one parent and other abnormalities in the second parent" do overlap. The first phrase asserts nothing about the qualities of the other parent and can very well subsume the second phrase. I admit that this might be too literal an interpretation of the subcolumn headings. However, there is another "overlap" within this table. A given proband could have a schizophrenic parent and also a schizophrenic grandparent and/or a schizophrenic aunt or uncle as well. Thus this particular proband could be placed in three separate columns (inflating the total percentages). The percentages in the last row of the table need not at all refer to different probands. Kallmann provides no information which would enable the reader to ascertain the degree of overlap.

7. In view of Hurst's comments on the reliability of the clinical records used by Kallmann (**1**, p. 405f) I wish to clarify one of my statements in this regard. My point referred to the possible change in observing and recording clinical symptoms as a result of Kraepelin's nosological contributions. One of Kraepelin's criteria for distinguishing between dementia praecox and manic-depressive psychosis was the presence of progressive deterioration in the patient. The introduction of this criterion may have led psychiatrists to look for and record those symptoms in the patient which seemed to be in line with the criterion. In pre-Kraepelinian days such a directed outlook was formally lacking and the determination of diagnoses could have been affected. Therefore an error of "indeterminate magnitude" is involved in the original clinical records of the probands—the initial set of observations with which Kallmann began his investigation.

It is indeed remarkable that Hurst is incorrect in *all* his allegations. I could only discuss a few of his objections in this brief note. The judgment I expressed in the review that the "Kallmann investigation . . . supplies no reliable information for assessing the genetic basis of schizophrenia" (**3**, p. 302) is emphasized by the weakness of Hurst's reply. One final remark—in the review I did not deny the possible genetic etiology of schizophrenia.

REFERENCES

1. Hurst, L. A. Genetics of schizophrenia: Reply to Pastore. *Psychol. Bull,*, 1951, **48**, 402–412.
2. Kallmann, F. J. *The genetics of schizophrenia.* New York: J. J. Augustin, 1938.
3. Pastore, N. The genetics of schizophrenia: A special review. *Psychol. Bull.*, 1949, **46**, 285–302.

Received February 18, 1952.

THE GENETICS OF SCHIZOPHRENIA: FURTHER REJOINDER TO PASTORE

LEWIS A. HURST

Alexandra Institution, Maitland, Cape Town, South Africa

In my original reply to Pastore my allusion to Haldane's and Hogben's view of Kallmann's work applied quite clearly to their reaction to his paper read at the Seventh International Congress of Genetics, Edinburgh, 1939 (my reference 5), the Proceedings of which I summarized for *Mental Hygiene* (my reference 1). Kallman's Edinburgh address concerned his American twin-family study to date, and not his earlier German study which constitutes the subject matter of his monograph *The Genetics of Schizophrenia* (my reference 3). There is thus no contradiction between my claim and Haldane's quoted assertion concerning the book.

So much for section 1 of Pastore's rejoinder. In replying to the other sections, I do not propose reiterating tabular and statistical details as our conflicting claims are now fairly and squarely before our readers for their arbitrament. I shall merely rehearse general principles and elaborate and clarify my original reply where this appears necessary.

The first part of Pastore's contention in section 2 has already been dealt with to my satisfaction. Regarding the latter portion concerning the genetics of tuberculosis and the hypothesis of gene-coupling between it and schizophrenia, it was not perhaps made sufficiently clear in my original reply that in the 1943 articles written in collaboration with Reisner (my references 10 and 11), Kallmann quite explicitly, on the basis of new evidence, modified his views as to *details of*, although *not the fact of* heredito-constitutional mechanisms in tuberculosis. The magnitude and significance of Kallmann and Reisner's contribution are forcibly brought home by the tribute of Barbara S. Burks on the occasion of the 1943 presentation of the findings.[1] One may add that the work of Kallmann's tuberculosis unit, exploring more detailed clinical and genetic aspects, has gone on unremittingly since 1943, and that a paper from Kallmann, Reisner, and Planansky is expected in the near future.

Section 3 of Pastore's rejoinder is the most veritable casuistry. I used the word "invariably" regarding numbers in subgroups for a frame of reference defined by Pastore's original criticisms, that is, where they are statistically relevant. The same considerations apply to cell entries. Pastore cannot claim that his play upon words here affects any scientific conclusion.

Sections 4 and 5 point, in my opinion, to a confusion between net and expectancy figures persisting in Pastore's mind. The fallacy of Pastore's criticism in his section 4 directed against table 58 in Kallmann's monograph, on the grounds of the sum of the independent probabilities of schizophrenia and schizoidia in the offspring of double schizophrenic parentage exceeding unity, is exposed by the following considerations. First, the table reveals quite clearly that of the 55 offspring of such unions, 32 only were schizophrenic or schizoid (16 of each). Then it must be borne in mind that the corrected rates of reference for schizophrenia and schizoidia here are 23.5 and 35 respectively, differing because of the different age distribution (apparent from the table) with reference to the manifestation period of the psychosis. To demand that the sum of these disparately derived expectancy figures should not exceed 100 per cent is ideologically unsound. An understanding of Weinberg's Abridged Method would have saved Pastore from this pitfall.

In section 6 Pastore objects to what he terms overlapping categories on the grounds of their inflating morbidity figures. He explains this by the example of a subject who may be entered in 3 columns in virtue of having not only one schizophrenic parent but also a schizophrenic grandparent and a schizophrenic aunt or uncle as well, and so counting as 3 instead of 1 when columns are totalled. What a quibble this is becomes apparent from scrutinizing Table 10 in Kallmann's monograph to which he specifically refers. It is found to consist of a series of 5

[1] BURKS, BARBARA S. Comment on paper by Kallmann and Reisner. *Amer. Rev. Tuberculosis*, **47**, 572.

tables separated by double lines (not columns of a single table) and there is no intention whatsoever of adding the totals. In cases where a subject genuinely plays a double or multiple role, Weinberg's proband and sibling method is specifically devised to correct the resulting bias. This method is fully described and frequently utilized in Kallmann's monograph.

One last word—I would ask the reader to turn his attention away from the minutiae of genetic interpretation to general outlines, which the hairsplitting of the present debate may have obscured. There is altogether too much of this sort of thing abroad today. Witness Bronson Price's attempt on the basis of anomalies in twin development to discredit twin studies in general and those of Kallmann in particular—although viewing these phenomena in perspective shows that they are of such rare occurrence as to render them incapable of significantly affecting the concordance rates, their differences and similarities, in representative samples of adequate size such as are employed by Kallmann. The long tradition of twin studies which springs from Galton is not likely to collapse in the face of such ill-founded attacks—witness Gedda's encyclopaedic work on twins (1950) and his recently founded twin journal.

Received June 23, 1952.

BOOK REVIEWS

HUMPHREY, GEORGE. *Thinking: An introduction to its experimental psychology*. New York: Wiley, 1951. Pp. xi+331. $4.50.

Five adjectives summarize the feelings I have about this volume: important, difficult, honest, scholarly, and historical. A defense of these adjectives should give a prospective reader some notion of the kind of book Professor Humphrey has written.

Important. This adjective is easy. The subject itself is important and no good summary has been available. This book is the best secondary source for research on thinking that you can buy today.

Difficult. Most of the book is difficult reading and some of it could even be called dull. This adjective gets less applicable as the text proceeds into more modern and more familiar territory. The most difficult chapters are those dealing with the Würzburg group and with Selz. These men found conscious processes that defied clear description. They invented or adapted German words to name them. Even in German these names connote more than they denote. Their concepts were painted on an intricate, highly elaborated, theoretical background that is completely inadequate and unfamiliar today. Translating from German to English and from 1910 to 1950 is a tremendous task in itself. It is not Humphrey's fault that the result is hard to read. The worst charge against him is that he devoted half the book to this translation.

Honest. The reading would have been easier if the author had been less honest in his reporting. The white lie, the begged question, the suggestive analogy, the over-generalization, and all the other dishonest devices that give respite to a weary reader are no part of this book. Each worker is reviewed systematically and usually is quoted at some length. Criticisms are given in the man's own terms and in the light of what he tried to do. No one is glibly denounced because he is dead and will not reply.

Scholarly. Patient scholarship shows through every page. Most psychologists are content with secondary sources and trust Boring, Woodworth, or Murphy for their opinions. We may scan some original sources and even read a few. But it is clear that Humphrey has studied the originals carefully until he understands what they say and what they do not say. The only place his diligent scholarship falters is in the decade from 1940 to 1950.

Historical. The organization is largely historical, beginning with associationism, adding the imageless thoughts and determining tendencies from Würzburg, then moving to Gestalt theories, motor theories, language, and generalization. Almost three-fourths of the references are dated prior to 1935. Rare and inaccessible sources are profusely quoted to preserve them for the future.

Because of his concern to keep the record straight and to deal honestly with each contributor, it is rather difficult to see what Humphrey's own position is. His earlier book, *Directed Thinking*, 1948, gives a clearer picture of Humphrey's general stand than does the present one. In many respects, the present volume is a teacher's manual to accompany the simpler version of 1948.

At the end of the book Humphrey outlines the present position of the experimental psychology of thinking. Fifty years of research are summarized in 16 statements. Thinking occurs when an organism meets, recognizes, and solves a problem. A problem is a situation that holds an organism from its goal. Thinking combines features of the problem which were originally discrete. It involves past experience. The form and method of the ingression of the past into the present is under dispute. Trial-and-error occurs. Motive is an aspect of thinking. Thinking is directed. The Würzburg group believed thought to be free from sensory content. They neglected the image which is a form of organization. Gestalt theorists have stressed productive as opposed to reproductive thinking. Thinking may be accompanied by changes in muscular tonus. Language cannot be equated with thinking. Generalization is a constant response to an invariable feature in a variable context. Images, action, speech, and concepts organize responses to a problem. Meaning is an artificial problem. At the rate of one such sentence for every three years of work, we have a long time to wait before we understand thinking. Perhaps the man who summarizes the first century will have at least one quantitative function worth drawing as a graph.

As the author says, "Fifty years' experiment on the psychology of thinking or reasoning have not brought us very far, but they have at least shown the kind of road which must be traversed." The prospect is not encouraging, but future travellers will surely agree that Professor Humphrey helped to show the way.

Massachusetts Institute of Technology. GEORGE A. MILLER.

KEMPTHORNE, O. *The design and analysis of experiments.* New York: Wiley, 1952. Pp. xi+631. $8.50.

This book is a worthy addition to Wiley's distinguished and growing list of titles in statistics. Although it is of little use to the beginning student, it can be profitably studied and used by psychological researchers at three different levels of statistical sophistication.

1. The experimenter who understands the meaning of least-squares methods and has some acquaintance with analysis of variance techniques will find a systematic and readable account of the tests with which he is familiar and can achieve a better appreciation of the standard tests and their possible variants even if he does not follow the mathematical details.

2. The student who commands only the elementary calculus can follow the general argument of the book (matrix arguments are usually set aside in separate sections) and will find a unified underlying treatment that should increase his understanding of the relations among the various tests.

3. For the reader who can handle matrix reasoning in addition to the calculus the book provides a satisfactory and eminently readable discussion of many

of the theorems underlying the statistics we use and the relations of these theorems to the interpretation of experimental data.

The reviewer particularly recommends an examination of Kempthorne's book to anyone who has attempted a systematic increase of his statistical prowess and found Fisher's works too elliptical and such books as Wilks' *Mathematical Statistics* too formidable mathematically.

The first five chapters are devoted to a brief discussion of scientific method with special consideration of the role of statistical interpretation and to a presentation and discussion of fundamental statistical ideas. More attention is given to problems of statistical estimation than is usually the case in the books on applied statistics widely used by psychologists. The teacher of advanced statistics should find these early chapters useful. The reviewer believes he can attest to the pedagogical soundness of the presentation given by Kempthorne, having worked out a similar approach (though neither so concise nor so complete as Kempthorne's) for use in his own classes.

In the later chapters a detailed discussion of an extremely large number of designs is given. These chapters should be of great value to the individual researcher and of perhaps even more value to those overworked members of departmental staffs whose duty it is to supervise the statistical aspects of graduate student research. The table of contents is detailed and descriptive so that it is easy to find information on any specific design.

In sum, this seems an excellent book for use either as text or as reference.

C. J. BURKE.

Indiana University.

JAHODA, MARIE, DEUTSCH, MORTON, COOK, STUART W., and others. *Research methods in social relations: With especial reference to prejudice. Part one: Basic processes. Part two: Selected techniques.* (2 vols.) New York: Dryden Press, 1951. Pp. x+759. $3.75 each vol., $6.00 the set.

These two volumes are intentionally different from each other in organization and level of discourse. The first of the volumes is the result of a cooperative effort among Jahoda, Deutsch, and Cook, and attempts to present an introductory, integrated account of "the considerations which enter into every step of the research process" (p. v) generally as well as those aspects of it specifically related to the area of social relations, with the illustrative material drawn from the area of prejudice where possible. The second volume is composed of a set of short chapters by various authors on methodological problems selected with the intent of supplementing in more technical detail some of the issues discussed in *Part One.* The audience at which the books are directed is broad: lay persons who are to act upon the findings of social scientists, students who are preparing to do social research, and social scientists who have not specialized in the field of social relations. These, then, are the knowledge areas and people aimed at by the works' authors. In evaluating how well they have hit their targets, neither the white bull's-eye disk nor the red, miss flag is appropriate.

Certainly, the approach is a laudable one; rather than the usual cold, abstract, machine-like picture of the research process, it is portrayed here as being actual behavior of real people and, therefore, beset by the same troubles as any like kind of social behavior. Certainly, the emphasis upon the necessity for coherent, explicit prior planning from hunch and hypothesis to data analysis and interpretation is commendable. Worth while, too, is the attention given to often neglected aspects of research such as fiscal and personnel administration. In general, then, where the books deal with what to do, how to do it, and when to do it, they are solidly based and make a real contribution. Nevertheless, some things about them are disturbing. Inasmuch as the volumes are so different in organization, these points will be taken up separately for each.

Part One. Unevenness in the level of discourse exists. Anyone who has to be told in Chapter Two that research is not a willy-nilly data collection affair is seldom ready to have Solomon's extension of control group design and the terms "analysis of variance" and "analysis of covariance" tossed at him in the next chapter. Examples of similar character may be found throughout the book and one suspects that sometimes the authors were considering their lay audience and at others the student or social scientist group. This flaw is not major in this volume; the authors' attitude and argument regarding values in research are, however.

Speaking to this point, it is easy to be misunderstood. Therefore, first, let it be asserted that few psychologists today would deny that values, personal and social, play a large part in their selection of a research problem area. Nor would it be denied by most that the social scientist is not removed from his social, institutional and cultural milieus; his rôle is that of the citizen-scientist. But it is a far cry from those assertions to the position that the scientist must act to produce social change, and yet this is the way the authors' argument runs. They appear to deny this by saying, "There is no valid scientific argument to compel or forbid [the social scientist] to encourage an [action] agency to apply his results" (p. 320). Yet this recognition is more verbal than real, since continual and heavy stress is laid upon research done for the purpose of promoting socially desirable objectives where "desirable" can only be defined by the researcher. And it is easy for the reader to infer that the authors as scientists (not citizens or even citizen-scientists) know scientifically what is good for society. This kind of interpenetration of personal values with data and action based upon those data is dangerous: it can lead to the espousal and support of an intellectual dictatorship potentially as bad (my valuation) as any other dictatorship; quieter, perhaps, but just as authoritarian.

As to the book's difficulties with theory, they may stem from the attitudes expressed about "action research." The following quotation appears to sum up the authors' own felt conflict between citizen and scientist: "One of the severest obstacles to an acceleration of the process of theory formation is the fact that those engaged in field research of the sort we have discussed in this book have not enough time to devote to it" (p. 335). And, even though a great deal of space is devoted to theory in the book, it remains a nebulous topic. As the book is read, one question keeps recurring: "All right. Theory is good and hypotheses are necessary. How are they constructed and used?" Then, too, one encounters statements like this: "If according to [the measuring] instruments, the prediction is not borne out, and if the value of the theory is beyond doubt (and this is a big "if" in the social sciences), [the investigator] will conclude that the instruments did not measure what they were designed to measure" (p. 110). Such a statement is naive. Is any theory ever beyond doubt even if it is not in the area of social science? Shouldn't a theory specify the empirical operations by means of which it may be tested?

Part Two. The most notable flaw in this volume of eleven chapters, each by different authors, is its extreme unevenness. Of course, unevenness is expected in a collection such as this, but the range here goes from an exhortatory chapter by Wormser and Selltiz on community self-surveys to a technical chapter on sample design by McCarthy. More editorial supervision would seem to have been indicated. (This last comment might also be applied to the organization of the two volumes taken together; there is a great deal of repetition of material.)

Value judgments and general problems of theory play a minor part in this volume; these are how- and what- and when-to-do chapters for the most part. Several of them are excellent references for the research worker. Notable among these are Kornhauser's guide to questionnaire and interview schedule construction, Proctor and Loomis' review of methods of analyzing sociometric data, and Stouffer's revision of Chapter I, Volume IV of *The American Soldier* series on scaling theory.

By way of summary, let it be said that considerable space has been spent on the negative aspects of the two books under consideration. This fact may leave the wrong impression. When these books deal with theory and values in science generally, it is easy to quarrel with them or find their approach inadequate, but when they deal with the so-called practical aspects of research, it must be said that they attain their goal of pulling together on an introductory level a large amount of valuable but otherwise scattered material on research in social relations.

J. C. GILCHRIST.

University of Wisconsin.

TILTON, J. W. *An educational psychology of learning.* New York: Macmillan, 1951. Pp. vii+248. $3.50.

This book treats the conditions of learning, particularly those that seem to have significance for education, from the viewpoint of field psychology. It should find a place as a basic or supplementary text in undergraduate courses. Its usefulness as such is enhanced by the inclusion of several chapters on educational measurement and the nature of intelligence that are largely independent of the author's systematic position. The book as a whole is too elementary for the graduate student of education, except for the nonspecialist in educational psychology.

Considerable space is given to the applications of field psychology to education. As has been true of others writing from this standpoint, the author appears to have been most successful in his handling of attention and perception as factors in learning, and least successful with motivation, particularly the acquisition of secondary motives. The subject of efficiency in learning also appears to be difficult to treat from the standpoint of field psychology.

In a work of this kind, it is always a question as to the extent to which educational ideas and practices are actually demanded by a particular system, or the extent to which they are as compatible with one system as another. In this respect, Tilton's book shows up rather well. In the first place, it is conservative in its claims, and in the second, it brings forth some ideas that stem from its systematic position. An example is the envisagement of the educative process as the "planned introduction of novelty in the experiences of the learner." Here the emphasis upon new patterning and new organizations of experience and reorganization demanded by new data is in the Gestalt tradition.

The book is scholarly and reserved. It is singularly free of extravagant and irritating claims that characterize much of current educational writing espousing field psychology. For example, the author does not see atomism as a necessary attribute of behaviorism or *S-R* theory, nor does he see trial-and-error and insightful learning as being mutually exclusive processes. Chiefly, he thinks a broader perspective is required since so much in human learning is poorly described in terms of reinforcement. Perhaps it will occur to many readers of his book that the reinforcement and field psychologists may sometimes work at different levels of description.

State University of Iowa. J. B. STROUD.

FILM REVIEWS[1]

Pronko, N. H., & Snyder, F. W. *Vision with spatial inversion.* 18 min.; silent; black and white. State College, Pa.: Psychological Cinema Register, 1951. Rental, $2.00 per day; sale, $37.00.

This film shows the changes that occurred in certain motor performances of a man during the course of an experiment in which inverting lenses were worn continuously for 30 days. The tasks which were systematically explored were: walking a complicated and very irregular chalk line, mirror-tracing of a star-shaped path, Minnesota Manipulation Test, Purdue Pegboard Test, card-sorting. For these the subject is shown performing *before* wearing the lenses, during his *first* wearing of the lenses, after wearing the lenses *24 days*, at the end of the experiment (*30 days*), and immediately after *removing* the lenses. Aside from stating when the performance occurred, there are few verbal titles and no interpretation.

The apparent objective of the film is to demonstrate that although inverting lenses cause a tremendous upset in the subject's motor behavior at first, it is possible in time to learn to react nearly as quickly and accurately with them on as without them. The qualitative evidence of this learning process is excellent, but it is regrettable that learning curves are not given to show the quantitative changes induced as a function of time.

The film would be appropriate in an elementary psychology course, as a supplement to the discussion of visual perception, presumably in connection with a consideration of the role learning plays. With considerable supplementary reading it might be helpful in an advanced course on perception or when dealing with the field of perception in an experimental course.

The chief criticism the present reviewer would make of the film is that it fails to *show* the psychological experience of the subject; it is stated that the lenses reverse the field of vision from top to bottom and from right to left, but this is no more than a textbook would state. There seems to be no reason why the film might not incorporate, for example, a picture of a room in the laboratory as it looks normally and as it appears when wearing the lenses.

A second omission is the failure to demonstrate the difference in learning tasks like card-sorting, which is definitely visually controlled, and the Minnesota Manipulation Test, which is almost exclusively kinesthetic. Presumably the learning curves for these two tasks were quite different, and undoubtedly they also showed a marked difference during the first lens trials, as compared to performance before putting on the lenses, but this is not brought out.

---

[1] *Editor's Note:* These reviews of films were prepared under the auspices of the Committee on Audio-Visual Aids of the American Psychological Association. Dr. A. A. Lumsdaine was chairman of the Committee at the time the reviews were written.

On the whole, this is a good job. It gives a feeling for an important perceptual problem and develops it clearly and understandably; the photography is good. Any theoretical interpretations will have to come from the instructor using the film; whether this is a constructive or a destructive criticism will, obviously, be a function of one's personal bias.

DOROTHEA E. JOHANNSEN.

Tufts College.

MILLER, NEAL E., & HUNT, GARDNER L. (assisted by Douglas Lawrence and Reign Hadsell). *Motivation and reward in learning.* 15 min.; sound; black and white. State College, Pa.: Psychological Cinema Register, 1948. Rental, $2.50 per day; sale, $60.00.

This film illustrates the significance of motivation and reward in the learning of albino rats. Its general thesis is that motivation activates the subject sufficiently to produce a wide range of responses. If one of these responses leads to a reward, i.e., a reduction of drive, the tendency to repeat that response will be increased. The wide range of irrelevant activities is gradually narrowed as the tendency to make the rewarded response is increased, until the subject is directly and efficiently performing the response that leads to the reward.

This thesis is demonstrated by observing the behavior of several rats in a modified version of a Skinner-box. The film starts with comparing a hungry and a satiated albino rat. The hungry rat is shown actively exploring the glass-front box. His responses are quick and varied. In the course of these activities he presses a bar which releases a pellet of food in the food cup. Gradually, his activities become confined to the vicinity of the bar and food cup. Eventually he learns to press the bar and eat the food with a minimum of wasted action.

A satiated rat, who has been placed in an identical apparatus, is first shown resting complacently. A different kind of motivation and reward is illustrated when a shock is presented to this satiated rat through the grid floor. His activity increases sharply and he quickly learns to depress the bar which, in this instance, ends the shock.

The general point is then made that *any* response which the subject is capable of making can be learned if it is followed by a reward. We are briefly shown several different rats in the same apparatus who have learned, respectively, to turn a wheel, bite a rubber tube, or fight with each other when these responses led to the cessation of shock. The film ends with a good summary of its main points.

It can be judged from the above synopsis that this film presents a simplified version of the role of motivation and reward in learning as conceived by S-R reinforcement theorists. This job it does very well. Observation of the rats in the film illustrates the point of view and procedure in a way that would be difficult to achieve through lectures or reading. It is perhaps best compared with a laboratory period or demonstration. Over these procedures it has the advantage (at least for students of elementary psychology) of eliminating the fumbling and presenting only significant details.

It is the opinion of this reviewer that the film would be a very useful adjunct to a lecture on the role of motivation and reward in learning for the introductory course and for undergraduate experimental psychology classes, particularly if the lecturer finds the S-R reinforcement approach palatable. But since it is a good demonstration of instrumental conditioning, it is likely that teachers with different theoretical inclinations can also find a place for it in their courses.

TRACY S. KENDLER.

New York University.

HAYES, K. J., & HAYES, C. *Vocalization and speech in chimpanzees.* 12 min.; sound; black and white. State College, Pa.: Psychological Cinema Register, 1950. Rental, $2.25 per day; sale, $50.00.

The film stresses two concepts. First, the chimpanzee will consistently utilize specific vocal sounds under specific forms of emotional stimulation. These specific vocal sounds are typical of the species in general and do not result from social imitation. The subject of this demonstration was isolated from others of the same species from birth. Sound samples of some of these typical vocalizations are presented:

The food bark during the period of anticipation of food.
The soft cry during low-level apprehension.
The high scream during high-level apprehension.
Laughter during tickling.

Second, the chimpanzee can be trained to vocalize certain sounds not naturally used; food objects served as motivation for the training and continuation of this behavior. Examples of these trained vocalizations are presented. These vocalizations have a rough imitative quality to the human sounds of mama, papa, and cup. The unnatural, trained sounds appear distinctively less smooth and effortless than normal spontaneous sounds. Examples of trained mouth movements are also supplied.

If the objectives of the film were to present clear-cut experimental results, these objectives have been impressively met. However, there is an absence of theoretical explanation, and of indication of the ramifications of the experimental results presented. Without such supporting interpretation, the film appears to have its greatest usefulness in advanced courses in general psychology, where theoretical aspects of the problems are known, or can be easily comprehended by the student when presented by the instructor.

The film appears to be average in photography, ingenuity of technique, and editing. The medium of the sound film has been excellently used to transmit the sounds of the chimpanzee voice.

EDWARD M. BENNETT.

Tufts College.

STONE, L. J. *When should grown-ups help?* 13 min., 16 min.; sound; black and white. New York: New York University Film Library. Rental, $4.00 per day; sale, $60.00.

STONE, L. J. *And then ice cream (children's meals).* 10 min., 16 min.; sound; black and white. New York: New York University Film Library. Rental, $4.00 per day; sale, $45.00.

These two films were produced at Vassar College in the series known as "Studies in Normal Personality Development." They are intended as class discussion material, for training of students in the techniques of observation with nursery-age children.

When Should Grown-Ups Help? presents four brief episodes in each of which a child is "helped" by an adult:

1. Andy is building a tall tower of blocks. In order to complete its top, he needs help. His teacher gives him the help. Then he is ready to start another tower-building project.
2. Penny is putting on her over-clothes. The adult (teacher or parent?) tries to help her, but Penny adamantly refuses, even though the adult keeps trying to help. Finally, Penny gets her clothing on without help.
3. Donald needs help in pounding a long nail into wood to join it with another piece of wood. He keeps hammering, but he makes no headway. He wants help but he gets none. He cannot complete his task.
4. Charlotte has caught her foot in the rope connecting her tricycle and cart. She twists and turns, while the teacher stands by without helping. Finally, Charlotte extricates herself.

And Then Ice Cream presents several cases of nursery-school children who are eating their school lunch. One eats slowly but methodically. Another rejects most of the meal. Ice cream dessert is considered an integral part of the meal, but it is reserved for the time *after* the food-eating period. Thus the ice cream dessert serves as a motivating factor in developing good eating habits.

These films may be considered most suitable for use with classes studying techniques of observation of nursery-school children. For such classes both the films present simple incidents which could stimulate much class discussion. In this connection, each film presents a set of discussion questions for consideration by the class. It is doubtful however, whether the films in their present form would be much suited for other purposes. They lack the wealth of clinical material which is found in many of the Vassar films, such as "This Is Robert" and "Frustration Play Techniques."

Photographically, the prints submitted for review appeared somewhat grainy, but this condition may be corrected in other prints. Technically, a device was employed which the present receiver found most distracting. Whenever the passage of time was represented, a caption, "TIME," separated the sequences. This device may have been an attempt to substitute for a mechanical "fade in–fade out" or "lap-dissolve," but the net result was not satisfactory, giving a rather "jumpy" sequence instead of smooth transitions.

A major fault of each film is the procedure used for review of the film content. The producers have simply repeated the film content verbatim. While this method of review has some merit under certain circumstances (e.g., if no one is available to rewind the film), the same result could be achieved by showing the film over again. Rather than review by repetition, a more appropriate procedure might have included summarization, or interpretation of some of the behavior demonstrated.

ELIAS KATZ.

Sonoma State Home.

VERPLANCK, W. S. and associates. *Testing intelligence with the Stanford-Binet.* 18 min.; sound; black and white. State College, Pa.: Psychological Cinema Register, 1950. Rental, $3.25 per day; sale, $75.00.

This is a simple presentation of the method of determining IQ with the Revised Stanford-Binet Scale.

An examiner administers some of the subtests to different children. Mental age (MA) and IQ are computed for children of similar chronological age (CA). A brief glimpse is afforded of some uses of intellectual evaluation in teaching, guidance, and counseling.

Since it is not intended as an intensive training film on the techniques of administration of the S-B Scale, the film would be most useful in elementary psychology and education courses. It would be especially helpful in courses where students had relatively little opportunity for practice in administering and scoring this test.

While the photography and the commentator's talk throughout are quite professional, the recording of the examiner's voice seems somewhat fuzzy in the early part of the film.

ELIAS KATZ.

Sonoma State Home.

BENNETT, A. E., & MCKEEVER, L. G. "*Antabuse*" *in the treatment of alcoholism.* 17 min.; silent; color. State College, Pa.: Psychological Cinema Register, 1950. Rental, $3.50 per day; sale, $85.00.

Interest in the use of Antabuse in the treatment of alcoholism is widespread. Consequently, any portrayal of the manner of its use and its effects are to be welcomed. The film shows the reaction which occurs when alcohol is drunk by a patient under treatment with Antabuse. By actually seeing what takes place in such circumstances, the viewer obtains an account of the reaction which would enable him to recognize it. In addition, information is given concerning the circumstances under which Antabuse therapy should and should not be undertaken. Roughly the latter half of the film follows a patient from the time of his admission to the hospital until his discharge. Some indication is given as to the place of Antabuse in the treatment armamentarium.

The film could be used successfully in a course in abnormal psychology or social problems. It could very well illustrate a lecture on the treatment of alcoholism. Although the film assumes some knowledge of medical terms, ignorance in this area is no handicap. Viewers should, however, have some orientation concerning alcoholism before seeing the film. The film might be shown to police officers who will have greater occasion to see the sufferer from an Antabuse reaction as use of the drug increases.

ALBERT D. ULLMAN.

Tufts College.

Vol. 49, No. 6 November, 1952

Psychological Bulletin

A CONSIDERATION OF SOME CONCEPTUAL TRENDS IN COMPARATIVE PSYCHOLOGY[1]

T. C. SCHNEIRLA
The American Museum of Natural History

HAVE WE A COMPARATIVE PSYCHOLOGY?

One might say that ideally comparative psychology exists when the comparative methods of science are applied to psychological problems. This must have been, in part at least, what Wundt had in mind in writing his *Vorlesungen* . . . (**181**). Potentially, such a program would appear to offer a good basis for bringing together and integrating too-divergent fields such as animal, child, and social psychology. Of course this type of approach has not disappeared altogether from psychology, for in recent years we have had an American edition of Werner's *Comparative Psychology of Mental Development* (**171**), utilizing animal, child, cultural, and social evidence in dealing with common psychological problems, and there is developmental psychology (**103**) which offers a somewhat limited and loose kind of alliance of child and animal study. However, the pre-contemporary and contemporary trends have led us from such integrative advances, for the most part, undoubtedly in some measure because of the difficult and tenuous nature of extrapolations among these disciplines. On the whole, comparative psychology is currently regarded as contained by animal psychology. But the digestion and assimilation have been uneasy and incomplete at best.

Although a sound use of the comparative method would be expected to unify and mutually strengthen all psychological fields having to do with problems of development, we have at the present time no articulate and integrated advance of the kind. One's strong impression is that no forward-looking program of the kind extends through animal psychology. Konrad Lorenz, the central figure in a vigorous contemporary group of European students of animal behavior, has commented as follows upon one symptom of our situation:

---

[1] The basis of a paper delivered by invitation at the symposium on Conceptual Trends in Psychology, annual meeting of the American Psychological Association in Chicago, September 1, 1951.

> Since the days of Charles Darwin the term "comparative" has assumed a very definite meaning. It indicates a certain rather complicated method of procedure which, by studying the similarities and dissimilarities of homologous characters of allied forms, simultaneously obtains indications as to the phyletic relationships of these forms of life and as to the historical origin of the homologous characters in question. I need not enlarge on the details of this method which is a commonplace to biologists and physiologists. We all know perfectly well what we mean by "comparative" anatomy, morphology, physiology, and so on. But it is all the more misleading if psychologists, who evidently are not familiar with what we mean when we speak of the comparative method, apply the same term in a very loose sense to all behaviour studies concerned with different forms of life. I must confess that I strongly resent it, not only from the terminological viewpoint, but also in the interests of the very hard-working and honest craft of really comparative investigators, when an American journal masquerades under the title of "comparative" psychology, although, to the best of my knowledge, no really comparative paper ever has been published in it (**87**, pp. 239f.).

If this somewhat abrupt comment is even partially correct, we have cause for concern. If the *Journal of Comparative and Physiological Psychology* does not live up to the senior term in its name, there may be cause for editorial study as to how far the journal *can* be representative under its present editorial program. But since this journal undoubtedly is representative of current developments to an appreciable extent, reactions such as that quoted above must be given very serious consideration.

Let us define the feasible limited objective: a comparative psychology of animal behavior, as the study of similarities and differences in adjustive capacities and personality among the types of living organisms (**128**). To what extent, if at all, do we have such a subject in American psychology? There appears to be no waning of interest in carrying out studies with animal subjects. In the decades from 1888 to 1940, in the annual total of articles appearing in 14 psychological journals, the percentage of articles dealing with animals below man has risen steadily from 3.5 to 15.2 per cent (**20**). Our deviations, whatever they may be, evidently cannot be attributed to any neglect to carry out investigations with lower animals.

The answer may lie in how and to what extent we study animals. A few years ago (**127**), I compared the *Journal of Animal Behavior* in its content from 1911 to 1918 with the *Journal of Comparative Psychology* in its content from 1938 to 1941, as to the kinds of animals used as subjects for the reported studies. A striking difference appeared, typified by the fact that while papers on invertebrate animals fell from 33 to 5 per cent, papers on the rat rose from 19 to 66 per cent. No recent change in this trend is apparent, for Beach (**13**) found later on that in 1946 and 1948, respectively, rats were the subject of 72 and 66 per cent of the papers in the *Journal of Comparative and Physiological Psychology*, all submammalian animals just 4 and 6 per cent. This trend toward concentrating on a single mammalian species as subject in the investigations of animal psychologists cannot be taken to mean that we have solved even the principal problems concerning inframammalian animals, notwithstanding the fine beginnings made earlier in this century by Yerkes, Jennings, Thorndike, and our other pioneers in the comparative field. Of course, in this space we cannot hope to offer any substantial analysis of the complex developments in biology and psychology, not to speak of the world in general, which must underlie the described symptoms. The practical consideration of paramount importance here is not how many types of animals we use in our investigations, but whether or not we use the comparative methods of science in working with them. Or perhaps the more primary question is whether we are even interested in such a goal.

Actually, we have developed no discipline of comparative psychology, either in investigation or in theory, of any substantial proportions. Just last year, a scientist in another field was prompted to write that "The life sciences and the sciences of the minds of both animals and men have been neglected too long" (**174**). Of course there has been a considerable amount of investigation of special problems such as sensory processes, neural mechanisms in learning, problem solving, and higher processes (**48**, **92**, **101**, **102**, **128**, **165**). The rat psychologists must not be held lightly, for they have accomplished many useful things (**27**). But their work has not favored, and in many ways has opposed, a trend toward a consistent, comparative study of behavioral adjustments and psychological capacities throughout the animal series. This is true notwithstanding the fact that the studies on animal capacities carried out by psychologists such as Thorndike, Yerkes, Watson, Lashley, Hull, and Tolman, to mention some contrasts, have occupied a basic and indispensable position in the development of modern American psychology.

However, the fact is not at all obscure that American psychologists have turned from comparative method to an anthropocentric interest in exploring specific problems, from more naturalistic procedures and attitudes to a technical, instrumental emphasis in pursuing these problems. Here we may consider a bit of human behavior which seems not unrepresentative of the general situation. Just last year at a conference, one of our well-known animal psychologists, speaking under the title "Levels of Integration along the Phylogenetic Scale—Learning Aspect" (**45**), explained that he would limit his discussion mainly to primates, because they are closest to man and well studied. Now primates are fine animals, interesting and important, but in a comparative sense they are not a very large part of the phyletic baseline. This attitude, we must admit, is typical, a sign of the times and no reflection on zeal or intelligence. In actuality, American psychology has turned sharply away from the beginnings of a comparative methodology perceptible earlier in this century.

Presumably, the broad responsibility of psychologists *as psychologists* concerns investigating all aspects and relations of behavior, personality, and social organization in the universe. Instead, with respect to animal investigations, the following anonymous statement seems to be the consensus:

> The principal if not exclusive problem of psychologists is learning about man and his psychological properties and putting this knowledge to good use. We learn most about man by studying man; however, lower animals closest to man can be used advantageously as subjects, because more conveniently obtained and manipulated. They are then used as substitutes for human subjects, and are considered equivalent to human subjects in the situation and problem of study.

This is assuredly a pivotal point in investigative and theoretical attitude, which certainly has led to much valuable knowledge for psychology in general. Some of the underlying assumptions, however, may not be as defensible as seems to be generally assumed. On the one hand, it seems manifest from a large body of sound experimental findings that, for many basic problems in reception and other areas in physiological psychology, rats may serve very well in place of human subjects. But many psychologists, for example Tolman (**161**) and Hull (**61**), largely take for granted the rat's equivalence to man as subject for many presumptive psychological problems, and especially for learning.

This conclusion is not foregone, as Köhler (**68**) has reminded us recently, although first admitting that "Even the rat may reveal great secrets." However, he then says, "But essential characteristics of man are barely discernible in the modest rodent; and what is barely discernible will easily be ignored. So . . . a selection of subjects for reasons of method gradually turned into a selection of special material evidence. . . . Knowledge has been and is being gained by separate rushes which aim at quite particular goals and for the time being at not much else. . . . Take the conditioned reflex, a notion about which scientific bias actually threatened to center." Köhler believes, however, that a "strong interest in the conditions of experimental proof," which he finds dominant among American psychologists, will overcome the hazards, and also that a trend toward a widened horizon of investigative and theoretical interest is apparent. Can we hope that such is the case?

It is very probable that a strong inclination at present discernible among general psychologists in the United States, to doubt the significance and even the relevance of animal evidence for human psychological theory (e.g., **7**), is in good part a reaction to the conditions referred to by Köhler in the above statements. These critics, from a strictly anthropocentric standpoint, do not seriously accept the post-Darwinian premise (**165**, **166**) that studying the animal "mind" is an unfulfilled responsibility of scientists. Not long ago Gordon Allport (**7**) gave explicit utterance to such a view, particularly on the ground that the gap in psychological capacities between man and his brute relatives is too great for any real help to be expected from infrahuman studies. Far from strongly negating such narrow views, most American animal psychologists at present seem to be *really* nonevolutionary minded, in the sense that they show no special zeal to find how man differs mentally from lower animals and vice versa, but rather focus strenuously on general problems without much attention to phyletic lines.

This practice seems dangerous to the depth and breadth of scientific behavior studies, even for the limited goal of understanding man. Even to justify using animals such as the rat, in a completely anthropocentric sense, as equivalent to man in given respects, there would have to be a substantial use of comparative techniques. How else could we demonstrate that whatever animals might be convenient to use were not too dissimilar mentally from man to meet the specific needs of the problem? Really, the entire discussion here seems rather schizoid to the present writer.

Criteria such as those now prevalent, emphasizing phyletic similarities and minimizing differences, may perhaps be taken to represent a contemporary form of anthropomorphism which subtly avoids mentalistic terminology but inevitably becomes loaded with equivalent implications. The point, of course, is not any serious encroachment of the specific superficialities of traditional mentalism, for the attitude of contemporary American animal psychologists is avowedly objective and *non*-mentalistic. Indeed, the specific superficialities of traditional mentalism are seldom noted. To the arguments of European phenomenalists such as Bierens de Haan (**16**) for a subjective and inferential description of animal "mind," we may oppose the comment of Boring (**19**) with respect to one contemporary objective system, that of Skinner (**143**): "The statements of functional dependency that result from such experimentation leave the concept of a controlling mind in the same limbo to which phlogiston has been consigned." Quite so! But, all the same, the mentalists may argue, with reason, that the operational and operant systems have made no great strides toward solving the problems of "mental organization" toward which the disowned concepts were directed. A useful hint for objective theorists may be found in Lashley's (**77**) remark that "the question 'What is the mental state of an animal?' means then: what is the level of organization of its activities." Statements such as this should be inspiring to those interested in the development of a comparative psychology (although they will not satisfy the minority of phenomenologists in our midst). For a discipline of serious contrasts is implied, with emphasis upon the significance of differences as well as similarities in capacities and adjustment patterns. But a scanning of the panorama of current American psychology suggests that we are not too close to this ideal. To investigate, perhaps the most efficient procedure is to undertake a brief comparative study of how some of our chief concepts are used contemporaneously.

"THE INNATE"

In a symposium on heredity and environment (**157**) published in the *Psychological Review* in 1947, Beach (**12**) says, "In agreeing to discuss any aspect of the so-called heredity-environment problem one naturally infers that the artificiality of the implied dichotomy is obvious to everyone." Unfortunately, this natural inference did not hold even in that symposium, for in summarizing Hunter was able to say: "It will come as a surprise to many psychologists that all five of the distinguished contributors to this symposium have emphasized the role of heredity in the determination of behavior" (**157**, p. 348). Now does "artificiality" in the first statement mean only that the dichotomy is "man-made," or that it is "man-made, and spurious"? If the latter alternative is the truer of the two, the surprise must be all the greater. It is the opinion of the writer, in his present somewhat uneasy role of trend-spotter, that a really weak dichotomy was accepted implicitly in the symposium. For the contributors agreed in general that the "anti-instinct writers" (**55, 71**) of the twenties went too far; then, in general, both by implication and by explicit ideology, they swung strongly toward the opposite extreme. We may inquire how far this procedure is a function of new evidence and of valid theoretical advances.

From the general literature one may gather that the aforementioned symposium did represent certain general trends toward a positivistic and recurrent emphasis upon behavior patterns as innate. "Instinct" certainly had the leading conceptual role. One of the contributors, in discussing whether so-called hereditary behavior is "innate or acquired," made the following statement:

> Adult rats, fully watered and fed, will hoard five to twenty pellets per day if placed in hoarding apparatus and left there. Since the rat only eats one or two pellets a day, this is plainly hoarding. And since it comes out spontaneously without training, it is plainly instinctive. Of course, rats may have learned to hoard early in infancy because of competition for food in their home cages, but hoarding is so universal in laboratory rats, always fully watered and fed, that such learning is unlikely (**100**, p. 336).

Certainly, many animal psychologists have not ceased to rely upon the criterion of universality to bolster an argument for pattern innateness, despite lessons from the past such as that conveyed by the "coenotrope" concept of Smith and Guthrie (**146**). For that matter, the main load of argument also frequently rests upon the criterion of early appearance, with experiential factors rejected out of hand if there are no obvious indications of them or if the experimenter has not introduced them programmatically (i.e., operationally). And in this exceedingly controversial field, teleological implications frequently lurk under the guise of supposedly functional terms. It may be suggested, for instance (**63**), that pulling-in or dragging-in behavior would be a proper operational substitute for the possibly improper term "hoarding," at least until we know more about what the rat's intentions are in such behavior. Retrieving behavior is also recognizable as a kind of "pulling-in" activity; and a common use of this term suggests the desirability of investigating possible relationships between these patterns.

There is no denying that much of our contemporary literature in this field is set along nativistic and preformistic lines. But also, as Anastasi and Foley (**8**) point out in a timely critique, psychologists often (and biologists not infrequently) "define heredity indirectly, vaguely, or inconsistently, especially when it comes to the domain of behavior phenomena." As these authors say, the definition of both heredity and environment, especially for the purpose of psychological investigation, is exceedingly difficult. They offer a provisional consideration of behavior etiology in terms of structural and functional factors instead of heredity-and-environment. How far such substitute dichotomies may be expected to reduce the fallacies in the hoary one of nature and nurture is not too clear. However, Anastasi and Foley believe the substitute should promote general understanding of statements such as Jennings' (**65**) that: "That which is directly inherited . . . is the set of genes, with the accompanying cytoplasm," in relation to Holt's (**55**) reminder that "No potential character ever is 'already contained' in anything . . . [so] that the applicability of the concept of heredity to behavior phenomena is indirect and remote." How well have the implications of statements such as these, based on sound theory in modern genetics, received consideration from animal psychologists and behavior students on either side of the Atlantic, in their investigations of so-called "innate behavior"?

The implication that "beyond the genes, which are the really inherited part of the organism, lies a problem of development," often seems elusive to behavior students. It seems unfortunate for progress along these lines that certain of our leading theorists have withdrawn somewhat from outpost positions once boldly taken. In the heredity-environment symposium, for example, Carmichael (**23**) discussed "The Growth of the Sensory Control of Behavior Before Birth" without reference to or use of his earlier significant theoretical contribution to this question (**22**), bearing on the development in utero of integrations of intrinsic and extrinsic factors influencing embryonic growth and activities. What intervening events in science have led to such changes is not clear. The significance of drug-inactivity studies with lower vertebrates (**21**, **164**) for this question now appears somewhat equivocal. The recent study of Fromme (**40**) indicates that a drugged condition of inactivation has different effects upon development of aquatic locomotion in frog tadpoles, depending upon whether the drug acts prior to stage 17 or during the critical stage 17–20 interval. When these matters have been clarified, there remains the question of how far evidence from such processes on the amphibian level can be extrapolated to the level of mammalian development. Although, as Sperry (**154**) says, "Holt (1931) and others earlier minimized neural organization available through maturation alone," it is equally dangerous to swing to the opposite extreme for vertebrates *in general* on the basis of results suggesting the role of matured neural paths for some of the *simpler* responses in the *lower* vertebrate classes.

Currently, emphasis is surely light on the need for systematic ontogenetic behavior studies and analysis of developmental stages considered as interrelated. Analysis is usually cross sectional, and is introduced by frequently arbitrary assumptions about procedures which supposedly rule out the determinative role of extrinsic factors during early stages. For example, the method of isolation, in which a laboratory animal is raised away from others of its kind, is very often adopted as a presumably rigid control for experiential factors. However, the risk of leaning upon this crudely teleological criterion is indicated by a consideration of how the individual animal itself, as a representative of "its own kind," may be a factor through processes based upon self-stimulative events in its own private activities. Consider a hypothesis, based on this point, with reference to the controversial problem of whether chicks recognize other chicks to any important extent on an innate basis (**58**, **110**). Since a segregated chick usually chirps when warmed, when fed, and in similar situations, a factor of learning may be introduced favoring its approach to other chicks as against ducklings, etc., in later "choice" tests. Controlling possibilities such as this represents a challenge to the theoretical vigilance and instrumental ingenuity of the investigator.

It is clear, as Leuba (**82**) states, that a systematic study of innate nature is needed. This should involve analytical approaches to the problem of behavior development in a variety of animals. Another need seems to be a new methodological equipment of improved controls on the experimental intervention of preconceived ideas of innate behavior patterning. It is suggested (**124**) that even an insect has an ontogeny which must be considered if we are to understand its full-blown behavior pattern.

As Hall (**42**) reminds us, the nature-nurture question is a pseudo-argument, which has not brought science closer to understanding the contributions of heredity to behavior patterns. The field of genetics in biology has concerned itself largely with morphology, whereas psychologists have studied behavior largely in pre-Mendelian terms. The growing interest in a joint discipline of "psycho-genetics," which Hall discusses, stresses the need for a *rapprochement* of these two fields in investigation. Such studies, performed on a systematic basis, should be very useful, as one source of evidence for a systematic theoretical examination of "innate nature." But investigations of hereditary-and-behavior correlations must be paralleled or extended by analytical studies of ontogenic stages in each animal type, if evidence from such research is to be properly evaluated and validly explained in theory.

There are very different attitudes toward the value for animal psychology of studies on morphological and physiological factors underlying behavior. Skinner (**31**) for instance maintains that a science of behavior may be developed without reference to neurology; Loucks (**88**) asserts on the other hand that neurology and neurophysiology cannot be excluded. Discussing the difficulties, Lashley (**76**, **77**) points to the fact that in some organisms where the relations are simple and direct (e.g., the earthworm) it has been possible to show correlations between specific structures and behavior, but in more complex reactions in which the influence of hereditary factors appears to be heavy (e.g., nesting in birds), clear correlative analyses have been baffled. Obviously, an even weightier challenge to understanding the role of structure in behavior is encountered in the processes of behavior modifiability. The challenge, however, is to improve theory, not to retreat with agility. How well such theory is improved may determine the relevance of evidence from psychogenetics for the problems of "innate behavior" and of ontogeny alike.

Now evolutionary theory and systematics in biology have placed a fundamental emphasis upon structural relationships among animal groups, implying some more or less important significance for adaptive adjustments in the respective groups. In the recent development of this field (**65**) this emphasis has been extended very actively to the consideration of behavioral factors in the adjustive repertoire of species. Animal psychologists, as Nissen (**106**) says, have found it difficult to see "how major psychological emergents are demonstrated to coincide with major and abrupt changes in structural characteristics differentiating larger taxonomic groups, especially the phyla." This difficulty certainly rests to a great extent in the lack of analytical comparative studies on patterns of behavior and their variations in related groups. The generalization that a direct involvement of structural factors in the functional determination of behavior patterns decreases phyletically from lower invertebrates to mammals (**92**) may be considered only a rather tenuous introduction to the necessary programmatic investigations.

If psychologists neglect to carry out such research systematically, others will attempt it, in their own ways. Biologists have recognized for some time that behavioral characteristics may be useful as clues to taxonomic affinities (**85**, **112**). The heuristic value of this fact has greatly stimulated interest of biologists in animal behavior. Thus Lorenz (**85**, **86**, **87**), leader of an enthusiastic contemporary group of European zoologists devoted to studying behavior, has been led to say that:

> . . . behavior patterns are not something which animals may do or not do, or do in different ways, according to requirements of the occasion, but something which animals of a given species have got, exactly in the same manner as they "have got" claws or teeth of a definite morphological structure. . . . From the recognition of this fact it is only a very short step to the systematic comparison of the innate behavior patterns characteristic of allied species. . . . [The term comparative] indicates a rather complicated method of procedures which, by studying the similarities and dissimilarities of homologous characters of allied forms, simultaneously obtains indications as to the phyletic relationships of these forms of life and as to the historical origin of the homologous characters in question (**87**, p. 238).

Such conclusions seem inevitable to those whose experience is bounded largely by specific behavior observations together with the realities of morphology. There is, however, much more to this than meets the eye, and perhaps psychologists should pay more careful attention to the validity of attempts to draw up principles of behavior in direct parallel with principles of morphology.

Not just because of its practical use to taxonomists, but more important, to observe and learn its nature, Lorenz urges behavior as an important object for biologists to study. He presents them with a rather neat, positivistic theoretical system (**85**, **86**) whereby behavior patterns are viewed as innately determined through intracentral processes in the nervous system, with specific stimulative effects acting as "releasers" to trigger them off. Lately, Lorenz (**87**) has begun to speak of natively determined patterns as "endogenous behavior." The investigations are essentially observational, qualitative, and relatively simple, mainly on birds although with studies on insects, fishes, and other forms also involved. The approach is avowedly comparative. However, since the theory is essentially preformistic, with an a priori emphasis upon the "native" aspects of behavior, a really comparative methodology of animal psychology has not yet appeared and may find difficulties in growing within this context. The proneness to impose a similar rigid nativistic ideology upon behavior in different and even widely different animals may be suggested by the final statements in a recent paper by Lorenz:

> . . . it is high time that social and group psychology began to occupy itself with the physiological side of behavior and more especially with the innate processes of which I spoke above. It is high time that the collective human intellect got some control on the necessary outlets for certain endogenously generated drives, for instance "aggression," and some knowledge of human innate releasing mechanisms, especially those activating aggression. Hitherto it is only demagogues who seem to have a certain working knowledge of these matters and who, by devising surprisingly simple "dummies,"[2] are able to elicit fighting responses in human beings with about the same predictability as Tinbergen does in sticklebacks (**87**, pp. 266 f.).

The Lorenz system has been criticized as "too simple." With all due credit to its desirable emphasis upon behavior study, the present writer wonders about the validity of its basic postulates.

That the Lorenz conceptual system is not without its appeal to American psychologists is indicated by the numerous references to it in the recent handbook (**155**) in which, however, no critical appraisal is undertaken. There are indications, it must be said, that a preoccupation with a presumed native predetermination of behavior may lead to overlooking necessary controls or significant aspects of evidence. Thus from an investigation frequently cited as evidence for Lorenzian principles, Tinbergen and Kuehnen (**160**) reported that the gaping (i.e., bill-opening) reactions of young thrushes are released initially by mechanical stimuli, but after the tenth day also by visual stimuli such as a black disc within a given size range in diameter. However, some observations by Lehrman (**81**) raise a serious question as to the validity of the Tinbergen-Kuehnen interpretation of the visual stage as natively determined. In the days immediately preceding that stage, the young nestling's eyes are likely to open during the latter part of the bill reaction (for example, while the parent bird is sitting on the side of the nest). A contiguity of stimuli is thereby effective with the critical response, which may introduce conditioning as a factor preparatory to the visual phase.

---

[2] Models, artifacts, used in the investigation of "releasers."

It is doubtful whether any methodology can be successfully comparative without a thoroughgoing concern for ascertaining the characteristics of ontogeny in each behavior system studied, as a step preliminary to the intellectual process of comparing behavior patterns and personality in different animals.

LEARNING

By and large, American animal psychologists have worked most intensively in the field of learning problems (**52, 101, 155**). Different methods have been used and differing theoretical viewpoints have arisen, represented currently by the stress of Hull's (**61, 62**) theory on "reinforcement processes" and by Tolman's (**161, 162**) on "field-cognition" processes in learning. At present, methods are more similar than previously; however, theoretical differences are still rather sharp between these two systems in particular (**79, 80, 152**). By and large, theoretical trends in this field are much too complex to receive any amount of treatment in the present context; hence we shall limit our discussion to some general comments which appear to be in order.

An observant bystander outside the immediate field of intensive investigation and research on learning readily gains the impression that those carrying on work in this field often are likely to grossly underestimate their own roles in the situation, as complexes of experimental variables. The crucial responsibility thus involved is indicated by Melton's (**98**) statement: "Stimulus response and field theorists frequently engage in the same types of investigations, with similar situations, but with such conditions imposed as to maximize the applicability of their own conceptual structure or minimize the applicability of the opponent's conceptual structure." It would seem that among the causes of contradictory results, in addition to the variations in investigative situation and poorly known conditions mentioned by Brogden (**155**), the experimenter himself must play a large and frequently uncontrolled role. His responsibility for narrowing down the scope of experimentation so that a limited and "unnatural" view of the phenomenon may be obtained in the end, is suggested by Brogden's (**155**) reminder that "The response has rarely been the independent variable," although it may be an important variable. Aiming for strict control, an experimenter laboring under the weight of a conventionalized theory may acquire a sort of "tunnel vision" which blocks from view aspects of the phenomenon which happen not to lend themselves to convenient study under his methods, or within his attitude. Though aiming to attain a *strict* experimental control, through losing the relation of trees-focused-upon to the forest he risks the loss of validity in the study of nature. Is it possible that the contemporary field of animal-learning investigation has become overspecialized along these lines?

The conditioning and the trial-and-error approaches to the study of learning both have existed in various forms for some time, often with applications to very different contexts (**52**, **166**). The influence of Pavlov's (**111**) and Thorndike's (**158**) findings and theory tended to be somewhat different in American psychology. Thus, after Watson's (**168**, **169**) inspired theorizing based on conditioning theory to a great extent, some investigators emphasize contiguity and reinforcement (**152**); after Thorndike's versatile studies of learning, with or without the influence of McDougall (**97**), other investigators emphasize the initiative, goal-seeking tendency, or cognition of the organism (**80**). Maier and Schneirla (**92**, **93**) distinguished between contiguity and selection as conditions making for qualitatively different processes in the learning adjustment; Skinner (**142**, **143**) and Hilgard and Marquis (**53**), from an operational point of view, distinguished between classical and instrumental patterns of conditioned learning. Culler and associates (**29**), in particular, reported results corroborating a qualitative distinction. To Kendler and Underwood (**67**), from the reinforcement standpoint, the results seemed insufficient for such a distinction; however, to Birch and Bitterman (**18**) the evidence seemed adequate, and further evidence was cited to support the position of Maier and Schneirla (**93**). Reinforcement-contiguity theorists (**62**, **152**) hold their lines to a concept of monomorphic learning without essential change. Tolman (**163**) has recently accepted the premise that qualitatively different learning processes exist; however, his distinguished processes *all* have the characteristics of cognitive mechanisms. It seems quite probable that both reinforcement and cognition formulae may be reducing the phenomena of learning to what may represent a *valid* phase, but a *different* phase of the problem range.

Thus far, the principal emphasis in studying learning has been upon the highly specialized investigation of a few mammalian types, without too much emphasis upon differences in patterns. The relatively few comparative studies attempted have brought interesting results. Thus, although a spider monkey investigated by Lashley (77) could not master a similarity relationship in a matching problem, this was easy for a chimpanzee. "An attempt to analyze the learning process here results in the discovery of apparently different processes." Rats and Formica ants in suitable scale replicas of the "same" maze pattern not only learned the situation in a strikingly different manner, but adapted to postlearning changes very differently, thus evidencing different types of learned organization (126). Although available theories present no very promising resources for dealing with comparative evidence, analytical studies of learning in very different animal types should reward us with important new evidence and fresh theoretical insight.

Because of limited evidence, conclusive phyletic comparisons concerning the properties of fixation (i.e., trace formation) or of organizational capacities in learning are premature. It is difficult to see how valid comparisons can be made, outlining the scope and nature of either similarities or differences, until the learning and learning-transfer processes have been exposed in each important animal type.

With respect to the feasibility of neurophysiological postulations in the study of animal learning, very different views exist. As a rule, neither Tolman nor Hull utilizes neurophysiological concepts in his theorizing. Skinner (**31**, **143**) maintains that a science of behavior can be built up without recourse to neural postulates. Lashley (**72**, **76**), on the other hand, consistently has utilized neural references in his learning studies, and recently Hebb (**47**) has offered a "redintegrative" theory of learning and perception, grounded in concepts of neural change with experience. It is to be hoped that we may soon discover a trend toward a genuine consideration of ontogeny and personality development in learning theory.

INTELLIGENCE

A direct challenge to the soundness of this concept is presented by Lashley (**77**) when, after stating that the spider repairing a web appears to exhibit an appreciation of spatial relationships of the same nature as concepts of spatial relationships basic to mammals, remarks that "The mechanisms of instinctive and intelligent behavior thus seem fundamentally the same." While no doubt overgenerous to the status of arthropod capacities, and although unsupported by analytical evidence, this comment nevertheless calls attention to our lack of systematic comparative studies of adaptive behavior on different levels. As Hebb (48) concludes, "We shall almost certainly find that the conception of a general level of intelligence is useless and that we shall have to learn how to define, not a single continuum, but a number of continua, requiring qualitative as well as quantitative analysis of behavior."

A stock definition of animal intelligence, "the capacity to meet new environmental conditions successfully," does not hold up well against the questions: "What conditions? How new? And how met?" To evaluate the process dependably, we must know more about the conditions of adjustment and the ontogenetic prerequisites than is usually known. Effective adjustment to an "emergency" offers an unstable criterion for intelligent behavior, for it is conceivable that different animals meet their respective crisis situations with equal degrees of success but on the basis of very different organizational processes.

"Complexity," often used as a criterion, is hardly standardized in meaning and is a weak pillar to support evaluations. As Nissen (**106**) states, although "the recurring term in phyletic comparisons is increasing complexity . . . [it is] of course too broad to be useful, unless the complexity is further specified." For example, in keeping with a complex morphological makeup, the starfish is capable of complex locomotive processes, in the sense of a very involved neural and stimulus-response repertoire, in which, however, a low-grade integration is predominant (**43**, **145**). It is doubtful that complexity may be taken as evidence for intelligence status, unless the concomitant qualitative characteristics are known. Formica ants can learn complex mazes, but the habits are organized in ways inferior to those of the rat, and their application of the habit to new situations is also inferior (**126**). In a "molecular" sense, trial-and-error solutions by higher mammals may be more *complex* than "insight" solutions.

One great difficulty is that the composite variable to be measured and compared, intelligence capacity, is necessarily influenced to variable extents by auxiliary factors such as sensory acuity, motor dexterity, and motivation. This entire "assisting caste," as Nissen (**106**) terms it, must be made functionally equivalent in the compared subjects, or otherwise controlled. It is of course an exceedingly difficult prospect to determine a weighting for the sensitivity factor, for instance, in animals as different as the rat and spider. For similar reasons, McBride and Hebb (**96**), after endeavoring to compare the intelligence of dolphins with that of higher vertebrates, reached the conclusion that at present no adequate method exists for this purpose. Not only do the auxiliary factors seem inextricably involved in the pattern, but the problem-solving process itself may be elusively different in different types of organism. These considerations suggest a need for really systematic efforts to solve the problem, not for giving up.

That results from original learning tests offer an elusive basis for comparing intelligence is indicated, for instance, by the difficulty Searle (**133**) experienced in trying to isolate a learning factor in analytical tests with "bright" and "dull" strains of rats. And in the "same" test situation, very different patterns of mastery may appear in initial learning (**94**). An extrapolation of human intelligence-test theory promises better results. A few beginnings have been made in devising batteries of tests in which, in various problem contexts, the ability to adapt a learned repertoire to new situations is measured systematically. Fischel (**37, 38**), in an interesting programmatic attempt to compare different animals in such terms, has tested reptiles, birds, dogs, and human subjects in a comparable situation. The subject is first permitted to master a characteristic adjustment to each of two lures of different attractiveness, then is tested in his ability to use these adjustments appropriately in a combined presentation.

Evaluating different ways of utilizing past experience is of course a most difficult assignment, as the history of the multiple-choice test has shown (**150, 184**). Using the discrimination-generalization test (cf. **114**) in a programmatic way, Harlow (**44**) and his collaborators have outlined the manner in which a monkey advances by stages of experience in the situation through successive "learning sets," from simple to complex and plastic reaction modes. Harlow criticizes the S-R theorists for their failure to investigate adequately the experience variable in its contributions to new situational adjustments. In most learning experiments, insufficient attention has been paid to this variable (**1, 25, 92**). Preparation is of course critical (**66, 109**), since experimenters working on both the animal (**17**) and the human levels (**33**) have obtained trial-and-error or insight solutions, depending upon the manner in which part-processes were contributed through controlled preliminary experience.

There is discernible in the literature a disposition to question the existence of an important qualitative difference between trial-and-error and insight solutions. Although this attitude is quite proper, and in accordance with Morgan's "canon," it may encourage too narrow a view in studying this crucial question. For instance, from his discrimination-generalization experiments, Harlow (**44**) is inclined to believe that insight is but a further stage of trial-and-error process in sequential learning, and not really qualitatively different from it. Prior to these Wisconsin experiments, however, Maier and Schneirla (**92**) questioned the adequacy of abstraction-generalization methods *by themselves* for differentiating "higher processes," on grounds which have not been controverted. These processes are undeniably related, but possible qualitative differences between them are not necessarily brought out by all types of test situation. The Wolfe and Spragg (**178**) investigation, designed to find whether Maier's (**89**) evidence for "combination of isolated experiences" could be reduced to the strict terms of contiguity-learning theory, did not succeed in this respect (**90**). In an answer to G. Allport's (**7**) challenge for clear evidence of higher processes in infrahuman animals, Seward (**134**) mustered strict criteria for insightful solutions which he found were met by Maier's recombination tests as well as by five other methods used to investigate such capacities in lower mammals.

Methodological and theoretical vigilance is of course essential to guard against the postulation of exaggerated solution capacities. Unfortunately, Morgan's canon is not accorded a basic place with equal seriousness by all students of animal behavior. Thus, in evaluating the remarkably manlike solutions achieved by monkeys in the Liège investigations, Lashley (**74**) concluded that the results were so influenced by uncontrolled variables (e.g., experimenter-derived cues) as to be without value for appraising the level of the solution processes involved (see also **56**). As another example, a sudden fall in a learning curve may indicate a solution process of high or low qualitative status; the shape of the curve is not a critical indication in itself, but the judgment must depend upon a detailed analysis of the experimental results (**92**).

Often the status of an observed situational adjustment seems to depend upon a sweeping characterization rather than upon a critical study of how the situational properties are met. Thus if maze learning is defined broadly along the lines of hormic theory as "goal response," then roundabout solutions may be attributed to all animals, from the earthworm to man, which have been able to learn some kind of maze (**15**). On the other hand, the concept of "Umweg" may acquire value for solution comparisons among widely different animals provided that tests and evaluation of results are guided by strict criteria (**38**). Yet again, the criteria of Umweg-solution are called into question by experiments such as those of Schiller (**120, 121**), from which roundabout solutions are reported for various inframammalian animals including cephalopods and minnows. In the same sense, the present criteria and meaning of "delayed-response" adjustments are really not adequate to appraise some reports of Thorpe (**159**), who describes cases in which solitary wasps maintained a route despite forced lateral deviations, and of Baerends (**10**), who cites the return of observed solitary wasps after intervals of time to provision appropriately each of two or three different burrows. Such developments remind us that some years ago the adequacy of the delayed-response method for calibrating higher adjustive capacities was challenged (**92**). One major difficulty is in controlling the extent to which critical cues are supplied by the animal or by the environmental situation. Far too little attention is being accorded to the question of evaluative criteria in discussing such problems. This condition seems not unrelated to a strong current tendency to merge superficially similar kinds of performance in very different animal types under the same phenomenal headings. Unclear comparisons with poorly devised concepts are considerably worse than no comparisons at all.

PERCEPTION

The application of this term to broadly different types of adjustment has increased to such an extent that clarification is urgently needed. On the one hand, experimental investigations have been directed at the analysis of perceptual phenomena (**47, 73**), and the continuity-discontinuity controversy (**75, 153**) is another expression of efforts to study the nature of situational meaning and organization in perceptual discriminations. On the other hand, use of the term on a vague qualitative basis in the general literature has increased rather than decreased.

There exists a widespread tendency to employ the word "perception" rather loosely to any case in which an animal senses an object or situation and reacts to it in a manner considered "adaptive" (e.g., **119, 156**). Whether the response is to be considered a reaction simply and directly elicited on an uncomplicated sensory-motor basis, or an appropriate reaction to the pattern of a recognized familiar object or situation, is very often very unclear. In the absence of specific evaluative criteria or even of a disposition to be critical in the matter, if such practices continue, the scientific role of this term in any conceptual sense may be reduced to that of a minus factor. Thus it is difficult to trace the specific implications of a statement such as Hartley's (**156**) for certain birds, that " . . . recognition of some enemies results from innate powers." Certain stimulative and behavioral relationships are described or implied in such cases, but the unknown status of the adjustment would seem to merit a more operational and less presumptive characterization.

Scattered experimental developments suggest that it is important to be vigilant for differences in the sensory-integrative processes underlying adjustments to situations by animals. For example, Hertz (**50, 51**) investigated the visual adjustments of bees to a variety of artifacts, and formulated a rather complex theory to account for the bee's ability to single out from the first those with greater brokenness of outline or greater internal complexity. Particular training was not needed; in fact, such response tendencies strongly opposed any training of positive reactions to the opposite visual condition. The investigations of Wolf (**176**) disclosed that such visually determined reactions in the bee are attributable to differences in flicker effect, and not to a visual pattern organization in the human sense so far as was discernible.

Unfortunately, the problem of "innate" contributions to perception has not been attacked very systematically and analytically on any subhuman level. The nature of differences between "those animals whose repertoire of perceptions is more or less limited to the relatively small number determined by innate organization . . . [and] those animals whose perceptual organizations are primarily acquired . . . ," as Nissen (**106**) suggests the distinction, is very poorly understood. What physical characteristics such as intensity, rate of movement, size, and so on, may produce adaptive reactions in animals without benefit of specific experience, is a subject that must be approached through methods appropriate to the animal type concerned. It is encouraging to note a disposition on the part of some psychologists to meet such problems frontally. Hebb (**47**), for one, has endeavored to develop a theory of perceptual development on the mammalian level, in the light of both native factors and of ontogenetic processes.

At the same time, less inductive approaches to these problems are in evidence. It is with problems concerning sensory apprehension and adjustment that Lorenz and his associates (**156**) have been especially involved. They have observed and classified for a considerable number of animals, particularly birds but also fishes and certain other lower vertebrates, the types of stimulus-effect ("releasers") which will produce given species-characteristic reactions. These reactions, as mentioned before, are considered innate and intraneurally determined, each with its corresponding extrinsic releaser, or trigger stimulus. The execution of the releasing-stimulus activity represents the end or goal toward which appetitive behavior is directed (**85**, **87**). The releasing stimulus (**156**) is typically named according to the observer's appraisal of the animal's (adaptive) behavioral relation to the situation of occurrence (e.g., an "enemy" valence). No particular differences are set up by these investigators between the postulated "innately given" qualities of objects or situations and the perceptual characteristics of learned, familiar situations in which the character of adjustment depends upon experience (**47**, **81**). Usually the releaser or the "sign stimulus" is described rather teleologically by the observer, whether it depends upon those features "most typical for the object" or "a configuration of features" (**11**). Some psychologists, troubled by the characteristic employment of observer-subjective naming devices instead of critically selected operational terms, may wonder as to our safeguards for reliability in weighting "innate" and "experience" variables, and for presuming to distinguish them in theory, when investigative methods are really cross sectional rather than ontogenetic. In view of our lack of adequate criteria for controlling such factors, the assurance of the outspoken Lorenz group is amazing.

Clarification is surely needed for the problem of "innately given" sensory adjustments to objects and situations, and their relation to the contribution of learning. How far are qualitative aspects (i.e., "meaning character") of the situation presumed to differ for the animal in innately controlled cases as against instances in which learning intervenes? To what extent and how does a pigeon "perceive *her young*," when upon their removal to a short distance she returns to sit upon the nest, while they shiver in full view? In preliminary investigations, Riess (**118**) found that female rats, after they had passed their early postweaning life in isolation without ever having handled or carried objects of any kind (subsisting entirely upon powdered food), in the parturition situation were very inferior in all respects to normal mothers, and lost all of their young promptly. What was the deficiency of these rats in early experience, as compared with normal rats? The hypothesis that the absent factor here is a perception of objects-to-be-approached-and-dealt-with (i.e., as "incentives") seems to merit testing. It appears that without the lacking factor a "maternal behavior pattern" cannot be formed.

The importance of experience for the development of visual perceptual adjustments in chimpanzees is indicated by Riesen's (**116**) investigation with chimpanzees raised in darkness from birth to 16 months of age. Since, in tests then carried out, objects evidently had no more meaning for these animals than for newborn chimpanzees, questions arise concerning the normal ontogeny of perception. (The possible contribution of previous perceptual training through other sensory modalities—e.g., somesthetic—as factors in re-education through intersensory transfer (**78**) will bear further consideration.) Such experiments indicate the development of a renewed interest in the ontogeny of individual perception in lower animals.

Investigations of the development of external drive controls in motivation contribute to an understanding of perceptual development. An early contribution was Woodworth's (**180**) concept of habits becoming transformed into drives through experience; a recent development has been the concept of "externalization of drives" (**9**). A predominant assumption is that motivational growth in mammals involves the specialization of the drive-behavior pattern with respect to particular learned situational adjustments. The characteristics of objects and situations which are connected with the relief of needs thus become focalized in perception, in a manner which much research on discriminative learning has been directed toward clarifying. The relationship of such processes to meeting the conditions of the environment is strikingly represented in the more rapid and effective acquisition of token-reward habit by chimpanzees (**177**) than by the cat (**147**), and its more plastic use by the former animal.

In the theory projected by Hebb (**47**) to account for ontogenetic perceptual development, the growth of perceptual patterns is traced for a mammalian animal from the naive stage through the first stages of learning. A developing action basis of expectancy and attention, a simple and molar one in this case, is followed through in an inductive procedure. This theory, which is characterized as a "redintegrative" developmental theory, appears to furnish a sound approach and good basis for further ontogenetic theory in child and animal psychology. Perception at any stage is regarded as involving an expectancy which is selective in function, and which thereby influences situational adjustments and further learning. What is learned at any stage depends on what can then be perceived, i.e., it is a function of how far the perceptual systems have advanced.

An inductive theory of perceptual development such as this calls attention to the problem of set or expectancy in situational adjustment and learning. It contains a further potentiality: namely, that relatively limited sets or adjustive patterns may represent early stages of more dynamic and versatile systems of expectancy, according to the animal's capacities for such advances. Is a conceptual clarification developing through such considerations—in that the term "set" may be applied more specifically to a functionally earlier, more limited (in selective scope) and stereotyped adjustive pattern? It is possible that, in contrast, a concept such as Maier's (**91**) "direction" represents the more dynamic expectancies of a later stage, through which the organism may overcome the bonds of specific contiguities more readily than before.

A groping for the relationship between perceptual and conceptual processes is currently apparent. It seems possible that perceptual systems, under appropriate conditions, may lead to conceptual systems. One interesting experimental development may be subject to reinterpretation in this light. After the Hayeses (**46**) had raised an infant chimpanzee under home conditions as closely similar as possible to those normal for a human infant, they reached the conclusion that this infant was inferior in no identifiable way to a human infant in its course of development until the language stage was reached in the latter. The principal psychological deficiency of chimpanzee ontogeny thus is attributed to the absence of a verbal language. On the other hand, it seems possible that preliminary stages of perceptual development, difficult to follow in observational studies, are partially similar in man and chimpanzee but also different in subtle qualitative ways. The hypothesis that the infant chimpanzee never reaches the threshold of perceptual-meaning-organization and nonlinguistic conception which is prerequisite for verbal-symbol mastery demands testing.

What is the relationship between perception and conceptual processes in mammalian ontogeny? Some writers seek an answer along the path of a "lumping" procedure. Thus Leeper (**80**) insists that the principal characteristics of "inductive concept formation are also to be found in discrimination learning and conditioning, and much of trial-and-error learning." These characteristics are given as differentiation and some generality of application. Whether this is to be considered a speculative exercise, or an introduction to a serious systematic reductive operation on the part of cognitive theory, is not altogether clear. It is often forgotten that in such exercises there is also a responsibility for trying to discern the significance of qualitative differences in the compared processes and capacities. It seems worth suggesting that a greater theoretical gain may lie in regarding these processes as successive stages in the development of an individual, i.e., in viewing conditioning as a process prerequisite to the more complex organizations of trial-and-error or selective learning, perception as a prerequisite to more highly involved functions of conceptual mastery, and the like. What is needed, perhaps, is a closer application of the comparative method to the tracing of stages in the development of personality on different animal levels.

A Comparative Social Psychology of Animals

Animal social investigations employing the comparative method are still uncommon, although there are indications of new developments both in research and in theory (**4, 125**). Psychologists have been less active here than biologists (**26, 132**). There are, however, a few indications of convergent developments in biology, psychology, and sociology (**3, 131**), pointing to further systematic investigation of social phenomena in animals. Advances are needed particularly in studying what differences may be significant for differentiating various types of animal societies, while at the same time their similarities are taken into account.

It is probable that the study of animal social behavior has resources which have not been brought to the investigative level attained by the study of human group dynamics (**4, 126**). Although much of the animal work is still descriptive, there are some detailed and systematic studies with implications for further research (**24, 30, 57, 123**).

The general conceptual trend is still influenced by intergroup similarities, much less by contrasts. An example is the concept of "dominance hierarchies," which has appealed to many laboratory investigators, including psychologists (**26, 95, 104, 122**). The "peck-right" concept has furnished a convenient way of describing the group behavior of vertebrate animals, and dominance-submission relations certainly have proved to be an outstanding aspect of group relations, as studied both in laboratory and field. It is now apparent, however, that this concept is somewhat like that of "extroversion-introversion" in the sense that it represents not just one social parameter such as aggression relations (**26, 95**), but a number of variables such as ascendance (**186**), prestige, and the like. Its main limitations may be summed up by saying that the dominance concept represents only one aspect of group behavior, which does not necessarily reveal the factors essential to group structure or may even reveal them in a negative way (**125, 129**).

The "dominance-hierarchy" concept, moreover, has influenced the attention and thought of students at the expense of concepts relating to the more positive and central aspects of group structure, so that the potentialities of concepts such as "cooperation" (**3**, **28**, **41**, **125**, **185**) seem largely unrealized. Although it is true of course that, as Zuckerman (**188**) has said, a comparative social psychology cannot exclude reference to the principle of dominance, it is now clear that unless dominance behavior is viewed in its relation to other social factors, investigation and theory in this field are certain to be seriously limited and distorted (**3**, **125**).

The usefulness of other concepts such as "cooperation" has been reduced greatly by tendencies to use them in a phyletically generalized way. Thus questions concerning the rise of gregariousness and the properties of group organization in different animal types have been approached only to a limited extent in research and theory. Investigations inspired by "innate" behavior theory (**156**) emphasize a descriptive type of study, concentrate on individual aspects of social behavior, and do not attack the problems of group structure in analytical ways. It should be noted that questions of group communicative behavior have been attacked on several levels (**14**, **39**, **107**, **115**, **125**, **187**), although mainly only in preliminary ways. However, an adequately comparative program of study of such phenomena is long overdue, particularly to clarify the relationships of concepts such as "sign," "signal," and "symbol," as well as the criteria of "language" (**14**, **115**), all of which appear to suffer from a heavy load of speculation and a minimum of systematic research. Research on questions concerning levels of phenotypic relationships through successive generations in lower animal groups, certainly basic to a needed re-evaluation of the broad problem of "culture" (**4**, **125**, **137**), should enlist the active attention of social psychologists, sociologists, and anthropologists alike.

It would seem that a sufficient number of objective studies of group behavior have been carried out (**24**, **28**, **30**, **57**, **123**, **132**) on widely different types of animals and groups, with prospects of others in a more intensive vein, to justify a new and more concentrated attack upon theoretical problems common to all group organizations. The basic biological aspects of such phenomena have been investigated theoretically by Allee (**3**, **5**) in particular. The central idea for a consistent theoretical statement as to the properties of "gregariousness" on different levels seems to be available in the potentialities of Wheeler's (**172**) concept of "trophallaxis" (i.e., food-exchange relationships). This concept has potentialities for a much broader extension in physiological psychology as a means of appraising the variant social impacts of fundamental biological processes in the growth and maintenance of different types of animal groups. A preliminary conceptualization of widely different types of group patterns, e.g., insect and higher mammalian structures, on this basis, has been offered by Schneirla (**125**, **129**) as the "bio-social" and the "psycho-social."

A limited but promising vigorous trend toward the study of animal group behavior has been apparent, and an inevitable enlargement of interest in the questions of the relationship of familial functions to the origin and continuance of larger groups is to be expected. There are a few beginnings in research on individual socialization in different animal types. A flourishing positivistic conceptualization of individual and group relationships in evolution, with the group characterized as "supraorganism" (**5**, **6**, **34**, **35**), has been found by some biologists to have heuristic value. On the other hand, it has been criticized (**108**, **125**) for having weaknesses common to general analogies, which in this case readily lead to serious misrepresentations of individual and group characteristics, and do not suggest or encourage comparative analysis.

ASPECTS OF METHOD AND ATTITUDE

Attitudes and practices concerning method are of course critical for the possibility of a genuine comparative basis for animal psychology. A few important points may be considered here. There appears to be a growing awareness that in the vast experimental literature on rat behavior, really very little psychological information about the pre-experimental antecedents and life of the experimental subjects is given as a rule. The subjects are usually regarded as "naive" and "inexperienced" if they have had no *specific* experience, to the experimenter's knowledge, in the *specific* kind of situation he is investigating. Of course, the literature for some time has not lacked reminders, and often strong reminders (**1**, **25**, **48**, **63**, **66**, **109**, **117**), of the fact that perceptual, emotional, motivational, and other influential aspects of the rat's personality must be considered and controlled as important experimental variables. Christie (**25**) calls attention to the probability that "these variations might have an appreciable effect upon the behavior of adult rats and might underlie some of the contradictory findings reported in the literature." Of this, it would seem, we can be positive.

This fact calls attention to the attitudes and predilections of the investigator himself as a most important variable in research on animal behavior. The reasons for a common lack of attention to the experience factor, and a general disposition to assume it to be homogeneous in the group and thus controlled, unfortunately cannot be considered simple and readily correctible. The symptoms center about a strong tendency to focus upon a specific problem without any adequate consideration of its broader naturalistic implications. There is in general too little interest in the ontogeny of behavior and personality development as important problems demanding investigation. Mainly for this reason, no doubt, a myopia is detectable for experiential factors which may influence behavior phenomena in somewhat obscure and indirect ways, and for their possible equivalence in function to "training" designedly introduced by the experimenter. As an example, Hudson (**59**) points out that:

> Observation of the behavior of an animal during the course of a typical conditioning experiment will often suggest that important aspects of the learning process are not available for scrutiny by that method. Resistance of the animal to being placed in the apparatus, for example, may develop long before the appearance of a conditioned motor response. This, as a relevant aspect of the learning, is sometimes overlooked.

Hudson emphasizes the significance of rapidly learned and very important perceptual reactions for the behavior of subjects in test situations. Cage life contributes psychological factors of this sort, however difficult they may be to identify. This point deserves renewed emphasis, to counteract the implicit assumption that, if subjects are "naive" for the *specific* problem, "training" then becomes a function of the experimenter's specific operations. This limitation may defeat the efforts of the most hard-working investigator for "rigid control."

As these considerations suggest, much contemporary work in animal psychology tends toward a segmental approach to the organism, through an intense concentration upon specific problems without a full appreciation of their behavioral setting. Thus, rigid control of a small-scale situation in the study of learning may mean a poor control in regard to the broad problem of personality development. A heavy burden is placed on the experimenter, proceeding operationally. Through his own subjective impressions and selective processes in evaluating the phenomenon, he decides what independent variables he shall measure and what "dependent variables" he shall control—devising situation, apparatus, and procedure accordingly. An experimenter, an S-R experimenter in particular, thus may not realize what he excludes when he selects for the animal what the animal is to learn. Planned along the line of "rigid" thought processes dictated by a particular theoretical structure, the experimenter's preconception of the phenomenon may be so imposed upon the investigative process that it becomes not a *test* of anything so much as it is a *demonstration* of that theory. As Harlow (**44**) has pointed out, S-R theorists do not really control the experience variable in learning studies, notwithstanding their stress upon a "historical" approach to learning. A similar error is made at times by cognitive psychologists in postulating systematic solution tendencies and insight in the animal's approach to a situation (**69**, **70**) without adequate attention to simpler alternative factors (**175**).

These are certainly not arguments against control in research on behavior problems; on the contrary, they suggest the need for basic *improvements* in control. Pre-experimental anecdotalists attributed higher mental processes to lower animals, or denied such processes, largely on the basis of stories. Their experimentally-minded successors, objectively oriented, may themselves indulge in a similar procedure more subtly derived from uncontrolled implicit processes in the experimenter as thinker and perceiver.

The expression *molar approach* is often used when the approach really is segmental. Research may be concentrated upon taking kinds of data which are partial, and thus may favor a really spurious answer to the problem. In the vast and devious literature on maze-learning in rats, how often would obscure results be clearer in their implications if the investigator had reported his data in terms of "kinds of errors" instead of (or in addition to) "error totals"? The experimenter's sagacity determines *what is significant* when he plans the investigation. He may be or feel under pressure to produce results, and thus may attempt a specialized study prematurely. As Sears (**31**) puts the matter, in appealing for really "molar" investigations in child study: "A kymograph cannot catch what is the essence of the child's conduct, whereas a human observer can." This is not a light or obvious point, as an uninformed bystander might think.

No argument is intended against quantitative methods in animal study, when the appropriate stage has been reached. The neglect of instrumental approaches at the required points in research may be serious for the validity of theory, as is suggested by Köhler's (**68**) statement that " . . . enthusiasm is no safe substitute for a high reliability of differences."

The compulsive advance of some psychologists toward phenomenalism, a development for which Tolman deserves a good measure of credit (**151**), must be recognized at least in part as one contemporary reaction to operationism, rigid use of quantitative methods, and deductivism in studying animal behavior. There has resulted an emphasis upon a strictly individualistic qualitative study of behavior which is often a poor substitute for a frank, pre-experimental "mentalistic" study of animals.[3] Such an approach may effectively exclude a consideration of behavior in a broadly naturalistic and social context, in the way that Sherif (**136, 137**) and Hebb (**47**), for example, undertake to study behavior.

An emphasis upon *similarities* in process and capacity is predominant in current animal psychology. For example, does the loose habit of applying the term "learning" to instances of temporarily changed behavior in lower invertebrate animals imply some essential equivalence to "learning" in mammals? Such phenomena in the low invertebrates exhibit features much like those of sensory adaptation and muscular fatigue in higher animals, occurrences not ordinarily spoken of as "learning." The terms perception and learning are endowed a priori with some vague essence of equivalence wherever used, or the characteristics of "cognition" are found generalizable throughout a range of processes from conditioning to reasoning. In effect, neither S-R nor cognitive theories have as yet shown their worth to differentiate the various kinds of adjustive and organizing process through the animal series, although some attempts have been made (**60, 61, 62, 161, 162**).

The concept "purpose," long a center of controversy in biology (**2, 32, 83**) and in psychology (**31, 55, 97**), has been receiving some study in behavior work. To this trend not only the behavioristic criticisms of Watson (**168**) but the more eclectic "purposive behaviorism" of Tolman (**161**) has contributed. Previously offered as representing impulsions toward goal or end situations through the entire scale of organisms, this concept is now more clearly seen as a confusion of "adaptive behavior" with a vaguely conceived mental process (**55, 85, 92, 129**). The term "purpose" itself is held to require the satisfaction of certain operationally definable criteria which thus far have validated the postulation of "anticipation," "expectancy," "goal-gradient," and similar higher-level processes only in higher mammals.

A clearer understanding of the relationship of quantitative methods to observational and essentially qualitative methods is very desirable. It may be contended that there need be no fundamental dissimilarity between observation and formal experimentation, considered as scientific disciplines (**36, 130**). To be sure, an important part of every behavior student's training should be learning how to recognize, and discount, an observational report which is merely anecdotal and hence unreliable. By the same token, he should also learn the clues to unreliability in experimental reports. "Hit-and-run" studies are found in both areas. Observational study, if performed systematically with an adequate methodology, can have the advantages of experimental control held as standard in laboratory work. Furthermore, a systematic observational investigation, through progressive stages of test and control, must inevitably lead into more formally devised laboratory experiments as an indispensable part of the program.

---

[3] It is interesting to note that, for different explicit reasons, S-R theorists (**61, 151**) and phenomenonalist theorists (**148, 149**) are inclined to exclude physiological evidence more or less completely. Notwithstanding this trend, physiological psychology continues to be a flourishing field of investigation (**155**).

These considerations suggest the advantages of correlative studies of animal behavior adjustments under natural and under man-devised (i.e., laboratory) conditions. A striking illustration is offered in the work of Hediger (**49**), who has compared with insight the adaptive behavior of the same animal types in the wild and under conditions of confinement in zoos. In general, however, animal psychology has been very deficient in attention to its naturalistic aspects. We should plan to learn how various types of animals may utilize their capacities (e. g., learning) in meeting the typical hazards and arduous trials of their habitats. By remedying this lack, we should gain a more adequate appreciation of the important types of animals as integrated, functioning organisms, and should be in a better position to select validly what is important for laboratory study.

THE DOCTRINE OF LEVELS

A concept of "levels" in capacities and organization follows inevitably from the theory of evolution. According to the levels doctrine, existing systems, both organic and inorganic, are viewed as comparable in their structural and functional characteristics (**105, 108, 125, 129**), and as capable of being ranked with respect to their degrees and kinds of similarities and differences in these respects. Such a concept offers the broad framework for a systematic study of phyletic similarities and differences which is urgently needed by animal psychology if it is to develop a comparative methodology.

One immediate gain from this viewpoint is vigilant attention to the validity of extrapolations from one system to another. How far, for example, may mechanical models be used in the study of organic and behavioral systems (**167, 173**)? The gains to be realized through a reference from mechanical calculating devices to the function of a mammalian brain may be elusive, since rough similarities may be stressed and crucial differences (e.g., the role of the human planner and operator of the machine) not really taken into account. As another example, evidence from the study of spinal-reflex systems has been extrapolated to the level of brain function, thereby imposing the pattern of a lower-level system upon a higher functional level in the organism. The investigations of animal psychologists, and of Lashley (**72, 76**) in particular, have gone far toward freeing psychology from theoretical error along this line.

A contemporary example of a reductionistic (**144**) theory is that of Lorenz (**85, 87**). This theory bases postulations concerning the innate control of general behavior patterns upon results such as v. Holst's (**156**), obtained from the study of *lower*-level neurophysiological functions in *lower*-vertebrate locomotion. The extrapolation here is from one type of pattern, locomotion (which might be considered a lower-level function in the individual), to organizations concerning the *whole* individual in action. Although it is a secondary matter here, the validity of the conclusions deduced from the lower-level function itself (i.e., of an endogenous, intraneural control of *patterns*) has been called into question (**84, 170**). Studies on the deficiencies of conditioned-response learning patterns in decorticate mammals (**139**) and the deficiencies of a reeducative recovery of locomotor control in spinal mammals as compared with intact individuals (**140**) serve to emphasize the qualitative differences of lower and higher part-processes in the organism.

The concept of levels has frequently come into discussion in the biological literature (**105, 108, 113**), less frequently in psychology (**135, 182**). Discussion of its potentialities for behavioral comparisons have been only introductory (**125, 129**) and occasional. The idea certainly is not a novel one with reference to behavioral capacities, since Aristotle described vegetative, sensitive, and rational souls for plants, animals, and man respectively, and both C. Lloyd Morgan (**99**) and Yerkes (**182**) utilized the essential principle of functional levels.

Certain generalizations may be derived from a preliminary comparative consideration of behavioral capacities in the principal types of living animals (**92, 129**). First of all, the relevance of evidence concerning morphological equipment may be considerably different according to the type of organism studied. Simply stated, in organisms considered lowest in a psychological sense, organic structure is found to have a rather immediate or "directly determining" efficacy for behavior, relatively unlike its significance for the pattern of behavior in higher organisms, in which varying stages of development and intervening part-processes must be considered. A second generalization is that any functional part-process (e.g., a sensitivity factor) may have different functional significance on different behavioral levels, according to the nature of the whole system within which it functions. Thus, on one level, knowledge about a particular functional component may be expected to have a different bearing upon whole-individual organization from that on other levels. This generalization sharply challenges the validity of extrapolations from level to level when the comparison is made in terms of similarities alone. In other words, extrapolations from level to level concerning similar part-processes have no necessary validity. If recapitulation doctrine has its limitations for modern biology (**54**), it has far less intrinsic validity for the theoretical psychologist.[4]

Consequently, the principles of behavior must be worked out carefully on an inductive basis for each important type of animal, as a foundation for carrying forward a valid system of interlevel comparisons. Comparisons based on similarities alone may be worthless, however striking they may be. Also, behavioral complexity as such offers no dependable basis for interlevel extrapolations, since complexity is found both in qualitatively high and in qualitatively low functional levels. The clear inference is that unless animal psychologists accept the task of making thoroughgoing functional studies of the principal and widely different types of living organisms, their generalizations will have validity only for the levels studied.

It is evident that knowledge and theory concerning phyletic positions and evolutionary background, as determined on biological grounds, must be treated as challenging and introductory by the animal psychologist, for whom the essential question concerns the degree of phyletic relationship or the lack of such relationships. For the psychologist, working contemporaneously, there remains the task of determining the significance of such evidence for the interpretation of behavioral capacities and organization in each type of living organism. Biologists consider existing animal forms as an incomplete series from which transitional forms have largely disappeared through natural selection (**141**). Psychologists must therefore study animal types as they find them, and rank them accordingly.

The levels viewpoint also emphasizes the dangers inherent in the teleological practice of naming behavior functions according to similar adaptive outcomes, such as food-getting, or protective behavior. Behavioral similarities alone, and similar adaptive results, offer deceptive clues to underlying processes in different animal forms. In terms of the levels concept, we become alert not only for evidence warranting the postulation of qualitative similarities, but also for qualitative differences at crucial turning points in the phyletic scale. A tendency to rely closely upon evidence concerning morphological turning points thus becomes educated through the consideration of behavioral qualifications. It is to be expected that difficulties will be experienced in relating such behavioral qualifications to structural turning points (**76, 77**), for at progressively higher psychological levels the relationship of specific structural equipment to behavior becomes increasingly devious.

---

[4] A limited heuristic value may be granted to Haeckel's law as applied to behavior, but no assured a priori place in theory.

Consequently, the meaning of the terms "part-process" (the molecular) and of more inclusive organization (the molar) must be worked out in a study of behavioral development and organization on each functional level. Then, by contrasting various levels, light may be thrown upon common problems in the field, such as learning. Premature comparisons, ruled by a preoccupation with similarities alone, cannot have much validity for a scientific, hence comparative, psychological theory.

Without doubt, a comparative study of behavioral development and capacities on different animal levels should stimulate investigations on the development of human capacities and personality, in ways perhaps otherwise unrealized. The answer to challenges of the relevance of animal-behavior evidence for human psychology is not that, in the absence of such evidence, irresponsible conclusions will be drawn and applied in any case, both in lay and in scientific circles. More significant is the greater breadth of view which can be obtained through comparative methods in studying the wide range of psychological problems.

This discussion may end with its initial statement concerning a defensible ideal in comparative psychology. One reason for encouraging such a discipline is to gain a valid grasp of man's similarities and differences with respect to lower animals, to replace the often irresponsible speculation which otherwise flourishes in general theory. A wider ideal is that the effective use of the comparative methodology of science in the basic fields of animal, child, and social study is certain to contribute to the conceptual development of all of these fields and thus benefit psychology as a whole.

REFERENCES

1. Adams, D. K. Experimental studies of adaptive behavior in cats. *Comp. Psychol. Monogr.*, 1929, **6**, 1–168.
2. Agar, W. E. *A contribution to the theory of the living organism.* Melbourne: Melbourne Univer. Press. 1943.
3. Allee, W. C. *The social life of animals.* New York: Norton, 1938.
4. Allee, W. C. Animal sociology.

Encycl. Britannica, 1948 Ed. Vol. **1**, pp. 971–972.

5. ALLEE, W. C. Extrapolation in comparative sociology. *Scientia*, 1949, **84**, 135–167.
6. ALLEE, W. C. *et al.* *Principles of animal ecology.* Philadelphia: Saunders, 1949.
7. ALLPORT, G. W. Scientific models and human morals. *Psychol. Rev.*, 1947, **54**, 182–192.
8. ANASTASI, ANNE, & FOLEY, J. P. A proposed reorientation in the heredity-environment controversy. *Psychol. Rev.*, 1948, **55**, 239–249.
9. ANDERSON, E. E. The externalization of drive. I. Theoretical considerations. *Psychol. Rev.*, 1941, **48**, 205–224.
10. BAERENDS, G. P. Fortpflanzungsverhalten und Orientierung der Grabwespe Ammophila campestris Jus. *Tijdschr. v. Entom.*, 1941, **84**, 71–275.
11. BAERENDS, G. P. Specializations in organs and movements with a releasing function. In *Physiological mechanisms in animal behaviour.* New York: Academic Press, 1950.
12. BEACH, F. A. Evolutionary changes in the physiological control of mating behavior in mammals. *Psychol. Rev.*, 1947, **54**, 297–315.
13. BEACH, F. A. The snark was a boojum. *Amer. Psychologist*, 1950, **5**, 115–124.
14. BIERENS DE HAAN, J. A. Animal language in its relation to that of man. *Biol. Rev.*, 1930, **4**, 249–268.
15. BIERENS DE HAAN, J. A. *Labyrinth und Umweg.* Leiden: Brill, 1937.
16. BIERENS DE HAAN, J. A. Animal psychology and the science of animal behaviour. *Behaviour*, 1947, **1**, 71–80.
17. BIRCH, H. G. The relation of previous experience to insightful problem-solving. *J. comp. Psychol.*, 1945, **38**, 367–383.
18. BIRCH, H. G., & BITTERMAN, M. E. Reinforcement and learning: The process of sensory integration. *Psychol. Rev.*, 1949, **56**, 292–308.
19. BORING, E. G. Review of *Current trends in psychology* (Wayne Dennis, Ed.). *Psychol. Bull.*, 1948, **45**, 75–84.
20. BRUNER, J. S., & ALLPORT, G. W. Fifty years of change in American psychology. *Psychol. Bull.*, 1940, **37**, 757–776.
21. CARMICHAEL, L. The development of behavior in vertebrates experimentally removed from the influence of external stimulation. *Psychol. Rev.*, 1926, **33**, 51–58; 1927, **34**, 34–47; 1928, **35**, 253–260.
22. CARMICHAEL, L. A re-evaluation of the concepts of maturation and learning as applied to the early development of behavior. *Psychol. Rev.*, 1936, **43**, 450–470.
23. CARMICHAEL, L. The growth of the sensory control of behavior before birth. *Psychol. Rev.*, 1947, **54**, 316–324.
24. CARPENTER, C. R. A field study of the behavior and social relations of howling monkeys (Allouatta palliata). *Comp. Psychol. Monogr.*, 1934, **10**, 1–168.
25. CHRISTIE, R. Experimental naïveté and experiential naïveté. *Psychol. Bull.*, 1951, **48**, 327–339.
26. COLLIAS, N. Aggressive behavior among vertebrate animals. *Physiol. Zool.*, 1944, **17**, 83–123.
27. CRANNELL, C. W. Are rat psychologists responsible for fission? *Amer. Psychologist*, 1947, **2**, 22–23.
28. CRAWFORD, M. P. The cooperative solving of problems by young chimpanzees. *Comp. Psychol. Monogr.*, 1937, **14**, 1–88.
29. CULLER, E. Recent advances in some concepts of conditioning. *Psychol. Rev.*, 1938, **45**, 134–153.
30. DARLING, F. F. *A herd of red deer.*

London: Oxford Univer. Press, 1937.

31. DENNIS, W. (Ed.) *Current trends in psychology*. Pittsburgh: Univer. Pittsburgh Press., 1947.
32. DRIESCH, H. *The science and philosophy of the organism*. London: Black, 1908.
33. DURKIN, J. E. Trial-and-error, gradual analysis, and sudden reorganization: An experimental study of problem solving. *Arch. Psychol.*, 1937, **30**, No. 210, 1–85.
34. EMERSON, A. E. Social coordination and the superorganism. *Amer. Midl. Nat.*, 1939, **21**, 182–209.
35. EMERSON, A. E. Basic comparisons of human and insect societies. In R. Redfield (Ed.), *Biological Symposia*, 1942, **8**, 163–176.
36. EMLEN, J. T. Techniques for observing bird behavior under natural conditions. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1103–1112.
37. FISCHEL, W. *Die höheren Leistungen der Wirbeltiergehirne*. Leipzig: Barth, 1948.
38. FISCHEL, W. *Die Seele des Hundes*. Berlin, Hamburg: Parey, 1950.
39. FRISCH, K. v. *Bees—Their vision, chemical senses, and language*. Ithaca: Cornell Univer. Press, 1951.
40. FROMME, A. An experimental study of the factors of maturation and practice in the behavioral development of the embryo of the frog, *Rana pipiens*. *Genet. Psychol. Monogr.*, 1941, **24**, 219–256.
41. GALT, W. The principle of cooperation in behavior. *Quart. Rev. Biol.*, 1940, **15**, 401–410.
42. HALL, C. S. The genetics of behavior. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 304–329.
43. HAMILTON, W. F. Coordination in the starfish. I, II, III. *J. Comp. Psychol.*, 1921, **1**, 473–488; **2**, 61–76; **3**, 81–94.
44. HARLOW, H. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
45. HARLOW, H. Levels of integration along the phylogenetic scale—learning aspect. In J. H. Rohrer and M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1950.
46. HAYES, K. J., & HAYES, CATHERINE. The intellectual development of a home-raised chimpanzee. *Proc. Amer. Philos. Soc.*, 1951, **95**, 105–109.
47. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
48. HEBB, D. O. Animal and physiological psychology. In C. P. Stone (Ed.), *Annual review of psychology*, Stanford, Calif.: Annual Reviews, Inc., 1950. Vol. II, 173–188.
49. HEDIGER, H. *Wild animals in captivity*. London: Butterworths, 1950.
50. HERTZ, M. Die Organisation des optischen Feldes bei der Biene. *Z. vergl. Physiol.*, 1931, **14**, 629–674.
51. HERTZ, M. Ueber figurale Intensitäten und Qualitäten in der optischen Wahrnehmung der Biene. *Biol. Zbl.*, 1933, **53**, 10–40.
52. HILGARD, E. R. *Theories of learning*. New York: Appleton-Century-Crofts, 1948.
53. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
54. HOLMES, S. J. *Organic form and related biological problems*. Berkeley: Univer. California Press, 1948.
55. HOLT, E. B. *Animal drive and the learning process*. New York: Holt, 1931.
56. HONIGMANN, H. The number conception in animal psychology. *Biol. Rev.*, 1942, **17**, 315–337.
57. HOWARD, W. E., & EMLEN, J. T. Intercovey social relationships in the valley quail. *Wilson Bull.*, 1942, **54**, 162–170.
58. HOWELLS, T. H., & VINE, D. O. The

innate differential in social learning. *J. abnorm. soc. Psychol.*, 1940, **35**, 537–548.

59. Hudson, B. B. One-trial learning in the domestic rat. *Genet. Psychol. Monogr.*, 1950, 41, 99–145.

60. Hull, C. L. Mind, mechanism, and adaptive behavior. *Psychol. Rev.*, 1937, **44**, 1–32.

61. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

62. Hull, C. L. The place of innate individual and species differences in a natural-science theory of behavior. *Psychol. Rev.*, 1945, **52**, 55–60.

63. Hunt, J. McV. The effect of infant feeding frustration upon adult hoarding in the albino rat. *J. abnorm. soc. Psychol.*, 1941, **36**, 338–360.

64. Jennings, H. S. *The biological basis of human nature.* New York: Norton, 1930.

65. Jepsen, G. L., Mayr, E., & Simpson, G. G. *Genetics, paleontology, and evolution.* Princeton: Princeton Univer. Press., 1949.

66. Karn, H. W., & Porter, J. M., Jr. The effects of certain pre-training procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.

67. Kendler, H. H., & Underwood, B. J. The role of reward in conditioning theory. *Psychol. Rev.*, 1948, **55**, 209–215.

68. Köhler, W. A perspective on American psychology. *Psychol. Rev.*, 1943, **50**, 77–79.

69. Krechevsky, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, **39**, 516–532.

70. Krechevsky, I. Hereditary nature of "hypotheses." *J. comp. Psychol.*, 1933, **16**, 99–116.

71. Kuo, Z. Y. A psychology without heredity. *Psychol. Rev.*, 1924, **31**, 427–448.

72. Lashley, K. S. *Brain mechanisms and intelligence.* Chicago: Univer. Chicago Press., 1929.

73. Lashley, K. S. The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.*, 1938, **18**, 123–193.

74. Lashley, K. S. Studies of Simian intelligence from the University of Liège. *Psychol. Bull.*, 1940, **37**, 237–248.

75. Lashley, K. S. An examination of the "continuity theory" as applied to discriminative learning. *J. gen. Psychol.*, 1942, **26**, 241–265.

76. Lashley, K. S. Structural variation in the nervous system in relation to behavior. *Psychol. Rev.*, 1947, **54**, 325–334.

77. Lashley, K. S. Persistent problems in the evolution of mind. *Quart. Rev. Biol.*, 1949, 24, 28–42.

78. Leeper, R. W. Study of a neglected portion of the field of learning—the development of sensory organization. *J. genet. Psychol.*, 1935, **48**, 41–75.

79. Leeper, R. W. Dr. Hull's *Principles of behavior*. *J. genet. Psychol.*, 1944, **65**, 3–52.

80. Leeper, R. W. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 730–757.

81. Lehrman, D. S. A critique of Lorenz's "objectivistic" theory of animal behavior. *Q. Rev. Biol.*, in press.

82. Leuba, C. The need for a systematic study of innate nature. *Psychol. Rev.*, 1940, **47**, 486–490.

83. Lillie, R. S. *General biology and philosophy of organism.* Chicago: Univer. of Chicago Press, 1945.

84. Lissmann, H. W. Proprioceptors. In *Physiological mechanisms in animal behaviour*. New York: Academic Press, 1950. Pp. 34–59.

85. Lorenz, K. Der Kumpan in der

Umwelt des Vogels. *J. f. Ornith.*, 1935, **83**, 137–213.

86. LORENZ, K. Ueber die Bildung des Instinktbegriffes. *Naturwiss.*, 1937, **25**, 290–330.

87. LORENZ, K. The comparative method in studying innate behaviour patterns. In *Physiological mechanisms in animal behaviour*. New York: Academic Press, 1950. Pp. 221–268.

88. LOUCKS, R. B. The contribution of physiological psychology. *Psychol. Rev.*, 1941, **48**, 105–126.

89. MAIER, N. R. F. Reasoning in white rats. *Comp. Psychol. Monogr.*, 1929, **6**, No. 3, 1–93.

90. MAIER, N. R. F. In defense of reasoning in rats: A reply. *J. comp. Psychol.*, 1935, **19**, 197–206.

91. MAIER, N. R. F. The behavior mechanisms concerned with problem solving. *Psychol. Rev.*, 1940, **47**, 43–58.

92. MAIER, N. R. F., & SCHNEIRLA, T. C. *Principles of animal psychology*. New York: McGraw-Hill, 1935.

93. MAIER, N. R. F., & SCHNEIRLA, T. C. Mechanisms in conditioning. *Psychol. Rev.*, 1942, **49**, 117–134.

94. MARKER, CHARLOTTE, & DOLGIN, M. Simple and complex bases for the solution of various tests in the Carmichael elevated-maze situation. *J. genet. Psychol.*, 1942, **60**, 259–269.

95. MASLOW, A. H. Dominance-quality and social behavior in infra-human primates. *J. soc. Psychol.*, 1936, **11**, 313–324.

96. MCBRIDE, A. F., & HEBB, D. O. Behavior of the captive bottlenose dolphin. *J. comp. physiol. Psychol.*, 1948, **41**, 111–123.

97. MCDOUGALL, W. *An introduction to social psychology*. (4th Ed.) Boston: Luce, 1911.

98. MELTON, A. W. Learning. In C. P. Stone (Ed.), *Annual review of psychology*. Stanford, Calif.: Annual Reviews, Inc., 1950. Pp. 9–30.

99. MORGAN, C. L. *An introduction to comparative psychology*. London: Scott, 1894.

100. MORGAN, C. T. The hoarding instinct. *Psychol. Rev.*, 1947, **54**, 335–341.

101. MOSS, F. A. (Ed.) *Comparative psychology*. New York: Prentice-Hall, 1945. (Rev. Ed., C. P. Stone, Ed., 1950.)

102. MUNN, N. L. *Animal psychology*. Boston: Houghton Mifflin, 1933.

103. MUNN, N. L. *Psychological development*. New York: Houghton Mifflin, 1938.

104. MURCHISON, C. The experimental measurement of a social hierarchy in Gallus domesticus: I. The direct identification and direct measurement of Social Reflex No. 1 and Social Reflex No. 2. *J. gen. Psychol.*, 1935, **12**, 3–39.

105. NEEDHAM, J. *The skeptical biologist*. London: Chatto, 1929.

106. NISSEN, H. W. Phylogenetic comparison. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 347–386.

107. NISSEN, H. W., & CRAWFORD, M. P. A preliminary study of food-sharing behavior in young chimpanzees. *J. comp. Psychol.*, 1936, **22**, 383–419.

108. NOVIKOFF, A. The concept of integrative levels in biology. *Science*, 1945, **101**, 209–215.

109. PATRICK, J. R., & LAUGHLIN, R. M. Is the wall-seeking tendency in the white rat an instinct? *J. genet. Psychol.*, 1934, **44**, 378–389.

110. PATTIE, F. A. Howells on the hereditary differential in learning—A criticism. *Psychol. Rev.*, 1946, **53**, 53–54.

111. PAVLOV, I. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.

112. PETRUNKEVITCH, A. The value of instinct as a taxonomic character

in spiders. *Biol. Bull.*, 1926, **50**, 427–432.

113. REDFIELD, R. (Ed.) *Levels of integration in biological and social systems.* Lancaster, Pa.: Cattell Press, 1942.
114. RÉVÉSZ, G. Experimental study in abstraction in monkeys. *J. comp. Psychol.*, 1925, **5**, 293–343.
115. RÉVÉSZ, G. The language of animals. *J. gen. Psychol.*, 1944, **30**, 117–147.
116. RIESEN, A. H. The development of visual perception in man and chimpanzee. *Science*, 1947, **106**, 107–108.
117. RIESS, B. F. A possible explanation of "freezing" behavior in rats. *Science*, 1945, **102**, 570–571.
118. RIESS, B. F. An experimental investigation of instincts in the white rat. The effect of restricted infantile environment on maternal behavior. *J. comp. physiol. Psychol.*, in press.
119. RUSSELL, E. S. *The behaviour of animals.* London: Arnold, 1934.
120. SCHILLER, P. H. Delayed response in the minnow. *J. comp. physiol. Psychol.*, 1948, **41**, 233–238.
121. SCHILLER, P. H. Delayed detour response in the octopus. *J. comp. physiol. Psychol.*, 1949, **42**, 220–225.
122. SCHJELDERUP-EBBE, T. Beiträge zur Sozialpsychologie des Haushuhns. *Z. Psychol.*, 1922, **88**, 225–252.
123. SCHNEIRLA, T. C. Further studies on the army-ant behavior pattern. Mass organization in the swarm-raiders. *J. comp. Psychol.*, 1940, **29**, 401–460.
124. SCHNEIRLA, T. C. Social organization in insects, as related to individual function. *Psychol. Rev.*, 1941, **48**, 465–486.
125. SCHNEIRLA, T. C. Problems in the biopsychology of social organization. *J. abnorm. soc. Psychol.*, 1946, **41**, 385–402.
126. SCHNEIRLA, T. C. Ant learning as a problem in comparative psychology. In P.L. Harriman *et al.* (Eds.), *Twentieth century psychology.* New York: Philosophical Library, 1946. Pp. 276–305.
127. SCHNEIRLA, T. C. Contemporary American animal psychology in perspective. In P. L. Harriman *et al.* (Eds.), *Twentieth century psychology* . New York: Philosophical Library, 1946. Pp. 306–316.
128. SCHNEIRLA, T. C. Psychology, comparative. *Encycl. Britannica*, 1948, **18**, 690–760.
129. SCHNEIRLA, T. C. Levels in the psychological capacities of animals. In R. W. Sellars *et al.* (Eds.), *Philosophy for the future.* New York: Macmillan, 1949. Pp. 243–286.
130. SCHNEIRLA, T. C. The relationship between observation and experimentation in the field study of behavior. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1022–1044.
131. SCOTT, J. P. Social behavior, organization, and leadership in a small flock of domestic sheep. *Comp. Psychol. Monogr.*, 1945, **18**, 1–29.
132. SCOTT, J. P. (Ed.) Symposium: Methodology and techniques for the study of animal societies. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1101–1122.
133. SEARLE, L. V. The organization of hereditary maze-brightness and maze-dullness. *Genet. Psychol. Monogr.*, 1949, **39**, 279–325.
134. SEWARD, J. P. The sign of a symbol: A reply to Professor Allport. *Psychol. Rev.*, 1948, **55**, 277–296.
135. SHEPARD, J. F. Types of learning in animals and man. *Psychol. Bull.*, 1914, **11**, 58. (Abstract)
136. SHERIF, M. Some methodological remarks related to experimentation in social psychology. *Int. J. Opin. Att. Res.*, 1947, **1**, 71–93.
137. SHERIF, M. *An outline of social psychology.* New York: Harper, 1948.
138. SHERRINGTON, C. S. *The integrative action of the nervous system.* New Haven: Yale Univer. Press, 1906.

139. SHURRAGER, P. S., & CULLER, E., Conditioning in the spinal dog. *J. exp. Psychol.*, 1940, **26**, 133–159.

140. SHURRAGER, P. S., & DYKMAN, R. A. Walking spinal carnivores. *J. comp. physiol. Psychol.*, 1951, **44**, 252–262.

141. SIMPSON, G. G. *The meaning of evolution.* New Haven: Yale Univer. Press, 1950.

142. SKINNER, B. F. Two types of conditioned reflex and a pseudo-type. *J. gen. Psychol.*, 1935, **12**, 66–77.

143. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

144. SLOANE, E. H. Reductionism. *Psychol. Rev.*, 1945, **52**, 214–223.

145. SMITH, J. E. Some observations on the nervous mechanisms underlying the behaviour of starfishes. In *Physiological mechanisms in animal behaviour.* New York: Academic Press, 1950. Pp. 196–220.

146. SMITH, S., & GUTHRIE, E. R. *General psychology in terms of behavior.* New York: Appleton, 1921.

147. SMITH, M. F. The establishment and extinction of the token-reward habit in the cat. *J. gen. Psychol.* 1939, **20**, 475–486.

148. SNYGG, D. The need for a phenomenological system of psychology. *Psychol. Rev.*, 1941, **48**, 404–423.

149. SNYGG, D., & COMBS, A. W. *Individual behavior.* New York: Harper, 1949.

150. SPENCE, K. W. The solution of multiple choice problems by chimpanzees. *Comp. Psychol. Monogr.*, 1939, **15**, 1–54.

151. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47–68.

152. SPENCE, K. W. Cognitive versus stimulus-response theories of learning. *Psychol. Rev.*, 1950, **57**, 159–172.

153. SPENCE, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 690–729.

154. SPERRY, R. W. Mechanisms of neural maturation. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 236–280.

155. STEVENS, S. S. (Ed.) *Handbook of experimental psychology.* New York: Wiley, 1951.

156. Symposia of the Society for Experimental Biology, No. IV. *Physiological mechanisms in animal behaviour.* New York: Academic Press, 1950.

157. Symposium on heredity and environment. *Psychol. Rev.*, 1947, **54**, 297–352.

158. THORNDIKE, E. L. *Animal intelligence.* New York: Macmillan, 1911.

159. THORPE, W. H. A note on detour experiments with Ammophila pubescens Curt (Hymenoptera; Sphecidae). *Behaviour*, 1950, **2**, 257–263.

160. TINBERGEN, N., & KUEHNEN, D. J. Ueber die auslösenden und richtungsgebenden Reizsituationen der Sperrbewegungen von jungen Drosseln (Turdus m. merula L. und T. e. ericetorum Turton). *Z. Tierpsychol.*, 1939, **3**, 37–60.

161. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

162. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948. **55**, 189–208.

163. TOLMAN, E. C. There is more than one kind of learning. *Psychol. Rev.* 1949, **56**, 144–156.

164. TRACY, H. C. The development of motility and behavior reactions in the toadfish (*Opsanus tau*). *J. comp. Neurol.*, 1926, **40**, 253–369.

165. WARDEN, C. J., WARNER, L., & T. N. JENKINS. *Comparative psychology.* Vol. 1. *Principles and*

methods. New York: Prentice-Hall, 1935.

166. Washburn, M. F. *The animal mind.* (4th Ed.) New York: Macmillan, 1936.

167. Waters, R. H. Mechanomorphism: A new term for an old mode of thought. *Psychol. Rev.*, 1948, **55**, 139–142.

168. Watson, J. B. *Behavior: An introduction to comparative psychology.* New York: Holt, 1914.

169. Watson, J. B. *Psychology from the standpoint of a behaviorist.* Philadelphia: Lippincott, 1919.

170. Weiss, P. Experimental analysis of co-ordination by the disarrangement of central-peripheral relations. In *Physiological mechanisms in animal behaviour.* New York: Academic Press, 1950. Pp. 92–111.

171. Werner, H. *Comparative psychology of mental development.* New York: Harper, 1940.

172. Wheeler, W. M. *The social insects.* New York: Harcourt, Brace, 1928.

173. Wiener, N. *Cybernetics.* New York: Wiley, 1949.

174. Wigner, E. P. The limits of science. *Proc. Amer. philos. Soc.*, 1950, **94**, 422–427.

175. Witkin, H. A. "Hypotheses" in rats: An experimental critique. III. Summary evaluation of the hypotheses concept. *Psychol. Rev.*, 1942, **49**, 541–568.

176. Wolf, E. Das Verhalten der Bienen gegenüber flimmernden Feldern und bewegten Objekten. *Z. vergl. Physiol.*, 1933, **20**, 151–161.

177. Wolfe, J. B. Effectiveness of token-rewards for chimpanzees. *Comp. Psychol. Monogr.*, 1936, **12**, 1–72.

178. Wolfe, J. B., & Spragg, S. D. S. Some experimental tests of "reasoning" in white rats. *J. comp. Psychol.*, 1934, **18**, 455–469.

179. Woodger, J. H. *Biological principles—A critical study.* New York: Harcourt Brace, 1929.

180. Woodworth, R. S. *Dynamic psychology.* New York: Columbia Univer. Press, 1918.

181. Wundt, W. *Vorlesungen über die Menschen- und Thierseele.* Leipzig: Vosz, 1863.

182. Yerkes, R. M. Animal psychology and criteria of the psychic. *J. philos. Psychol. sci. Math.*, 1905, **2**, 141–149.

183. Yerkes, R. M. *The dancing mouse.* Leipzig: Vosz, New York: Macmillan, 1907.

184. Yerkes, R. M. The mental life of monkeys and apes: A study of ideational behavior. *Behav. Monogr.*, 1916, **3**, Ser. No. 12, 1–145.

185. Yerkes, R. M. Genetic aspects of grooming, a socially important primate behavior pattern. *J. soc. Psychol.*, 1933, **4**, 3–25.

186. Yerkes, R. M. Social dominance and sexual status in the chimpanzee. *Quart. Rev. Biol.*, 1939, **14**, 115–136.

187. Yerkes, R. M., & Yerkes, Ada W. *The great apes.* New Haven: Yale Univer. Press, 1929.

188. Zuckerman, S. *The social life of monkeys and apes.* London: Kegan, Paul, 1932.

Received February 8, 1952.

Vol. 49, No. 6 Psychological Bulletin November, 1952

THE SOCIAL BEHAVIOR OF VERTEBRATES: A REVIEW OF THE LITERATURE (1939–1950)

WENDELL I. SMITH
Bucknell University
AND
SHERMAN ROSS
University of Maryland

An experimental approach to some of the biosocial problems of human social interaction is frequently made through studies on lower animal forms. Beach (**28**) states that, "We are just beginning to appreciate the usefulness of a comparative study of social behavior. The findings . . . point to the potential advantage of using a variety of animal species in our investigations of interactions between members of a social group" (p. 123). Furthermore, he stated that unless experimentalists escaped the spell that *Rattus Norvegicus* was casting over them, even the experimentalists themselves were in danger of extinction (p. 117). Beach reached this conclusion from an analysis of the studies reported in all odd-numbered volumes of the *Journal of Animal Behavior* and the *Journal of Comparative and Physiological Psychology*. Beach's thesis in this paper is that a comparative study of social behavior based on a variety of species would be a most valuable scientific enterprise.

That there has been a dearth of studies of social behavior of animals might seem to be the case if one referred to the chapter by Nissen (**168**) in the recently published *Handbook of Experimental Psychology*. On the other hand, reference to chapters by Nissen (**169**) and Collias (**72**) in Stone's *Comparative Psychology* would indicate that the study of social behavior of animals has not vanished. The reported studies, however, are scattered in many journals, and most certainly are not (in full agreement with Beach's analysis) concentrated in the main in psychological journals.

The purpose of this paper is to bring together the scattered literature in review form so that ready reference to the rather extensive literature is available to the reader. Crawford (**77**) has reviewed the literature to 1939, and the present writers have covered the period from 1939 to 1950, following the general organization used by Crawford.

There are a number of references covering a wider range of topics than will be considered here, as well as several references which are reviews of particular aspects of social behavior. Allee *et al.* (**16**) have summarized much of the literature on animal ecology, and this text is of value for the chapter on animal aggregations. The principles of cooperation and disoperation are proposed as explanatory concepts of sociality. This general view of Allee's is presented in other, more limited discussions (**4**, **5**, **7**, **8**, **9**,) which emphasize the survival values of social behavior. Other reviews and discussions of social behavior may be found in Tinklepaugh (**234**), Darling (**87**), Lack (**146**), Harding (**128**), Faure (**98**), Walker (**239**), Cattell (**64**), Redfield (**178**), and Bruce (**52**).

An excellent monograph on the methodology for studying animal societies has been edited by Scott (**212**). The papers by Carpenter (**63**) and Schneirla (**193**) are useful for those who are interested in applying objective methods to the study of behavior under field conditions. These two writers present a very convincing discussion of the applicability of techniques customarily associated with the laboratory to animal groups living in their natural habitat. Carpenter's success in this area (**57**, **58**, **60**) is well known. Scott's (**208**) monograph on sheep and his publications on mice (**205**), goats (**206**), and dogs and wolves (**213**, **214**) represent a more limited application of objective methods to the study of social behavior in "natural" groups. Scott (**216**) has also been successful in gathering significant data on the social behavior of three little known species of grouse. Reported studies of various groups of vertebrates living in seminatural habitats in zoological parks indicate the value to a comparative social psychology of studying a range of species. For summaries of this work, see Collias (**69**, **70**, **71**, **72**). Consideration of other methodological problems is given in Fuhr (**107**) and Hebb (**133**, **134**).

Haskell (**129**) has presented criteria for a natural classification of all societies, including that of man. The classification is determined, in part, by a property called "coaction," a characteristic of a situation in which the "stronger individuals either increase or decrease this (life) process' normal rate in weaker ones, or vice versa, or both" (p. 187). In a related discussion, Essen (**97**) analyzes the interpersonal behavior of animals and concludes that the interactive evaluation is primarily dependent upon physiological and learning factors. Following this general line of a quantitative evaluation of interaction, Dice (**89**) suggests the use of Forbes's coefficient of association to provide a measure of the amount of association of one species with another.

AGGREGATIONAL BEHAVIOR

The best source of information on aggregations or "contagious distributions" may be found in the work of Allee and his collaborators (**3**, **4**, **5**, **8**, **10**, **16**). The various factors responsible for aggregations and one theoretical explanation of them can be found in Allee *et al.* (**16**).

Other theoretical discussions of aggregation have been presented by Fraenkel and Gunn (**103**), and Rabaud (**176**).

Fish. Practically all of the experimental work in the past decade on the topic of aggregational behavior in vertebrates seems to have been limited to fish. Allee *et al.* (**15**) have found a favorable effect on the growth of goldfish which are raised in water in which other individuals of the same species have resided. This effect would be predicted by the principle of coaction (**129**), which has been supported primarily by the findings of Langlois (**147**) on bass. Shlaifer (**223, 224**) has reported that the goldfish has a lower rate of locomotor activity and oxygen consumption when grouped than when isolated. This "group effect" is said to be dependent on visual responses to other goldfish or "goldfish-like forms." Breder and Nigrelli (**45**) have confirmed an earlier finding that the swimming speed or locomotor activity of fishes is related to the size of the aquarium and the number of fish present. Data were obtained in each of three dimensions by ruling coordinates on the glass walls. Breder (**44**) discusses measures of quantifying the degree of the aggregating tendency. By using models of various sizes and live fish of matching colors, several interesting results were obtained. Targets of a particular size led to approach, but if a larger target was presented, retreat occurred. Groups of like color seemed to be more cohesive than groups containing fish unlike in color. Isolation from the group for several months modified the aggregating pattern and the isolates were sensitive to previously ineffective stimulation.

Schlaifer and Breder (**225**) report a tendency for tarpon to rise to breathe when other individuals rise. Using models, they were able to induce "imitative" rises in tarpon. These rises are said to occur only when there is a depletion of oxygen in the swim bladder. In a later study, Breder (**43**) reports no increase in the number of rises with an increase in temperature as had been found in small tarpon. Furthermore, nearly every rise in the mature tarpon was individual and not imitative as in the small fish.

Size plays an important part in the isolation of minnows from a well-habituated minnow shoal according to Berwein (**35**). Minnows smaller or larger than the shoal by approximately twelve millimeters are isolates. This isolation is not affected by temperature or extirpation of the cerebrum.

Noble and Curtis (**171**) report evidence that the breeding colors of fish serve as cues to attract the same species. Fish which develop red breeding colors are attracted by red in early life while those which develop black colors are more attracted by black. The association with parents appears to strengthen the native tendency to approach disks of a particular color since isolated fish respond to them relatively slowly in comparison with fish raised with the parents. The three-spined stickleback has been studied extensively by Tinbergen and his co-workers (**175, 232**) and some of his results confirm the findings of Noble and Curtis. Tinbergen (**232**) prefers to use the releaser concept as a working hypothesis, which he describes in part as follows:

> Innate behavior is never a reaction to an environmental situation as a whole, but only to a few parts of it. Other parts—although they may be received by the animal's sense organs—are entirely ignored . . . (p. 6).
>
> . . . only innate behavior is dependent on sign stimuli . . . every unconditioned reaction is dependent on its own special sign stimulus and different reactions respond to different sign stimuli. This shows that the release of every unconditioned reaction is dependent on a special central nervous mechanism, which is called the "innate releasing mechanism" (p. 7).

Stickleback are said to display releasers of form, color, and movement (**175, 231, 232**). Using various models of different forms, it was found that the red belly acts as a visual releaser. The type of movement exhibited also serves as a releaser. A model placed in the aquarium with head down elicited fighting, while the same model with head up elicited following behavior from other fish. The swollen abdomen of the pregnant female serves as a strong releaser for mating behavior. Thus, aggregations in this fish are frequently dependent upon the response of one fish to the cues provided by another.

Fisher (**100**) investigated the effect of temperature on trout and salmon, using a channel aquarium in which the temperature varied from 0°C. at one end to 30°C. at the other end. The fish, which had been acclimatized to 4°C. congregated at two points, the trout at 10°C. and the salmon at 14°C.

Amphibia. The reaction of tadpoles to light has been studied by Folger (**102**). Aggregations formed at the lighted end of the jar. The number of animals aggregating differed from group to group, and from day to day within the same group. Several chemicals affect this reaction to light.

A field study of the toad by Freisling (**106**) indicates that the geographic distribution of this animal is largely dependent upon a thigmotactic response. The skin of the thigh and of the abdomen is most sensitive to desiccation, and the toad adjusts, within temperature limits, to climatic differences in terms of the moisture content of the substratum against which the sensitive skin is pressed.

It is characteristic of the studies cited that phylogenesis rather than ontogenesis has dictated the hypotheses which have been tested. Schneirla (**194**) has questioned the wisdom of failing to investigate the ontogenetic development of behavior patterns in lower animals because of the belief that the behavior is natively determined. For example, Schneirla takes issue with the Lorenzian concept (**149, 150, 151, 152**) of a "surrogate neural center," viz., the releaser for a specific behavior pattern, and presents an analysis of the so-called released behavior in terms of simple conditioning. Apparent universality of a behavior pattern is said to be in part a function of a relatively common environment containing the elements requisite to conditioning.

This theoretical approach to the study of aggregations would perhaps serve the end of pointing up the role of early learning in highly stereotyped behavior.

SOCIAL FACILITATION

In Crawford's review (**77**) of the literature, the term social facilitation was used to refer to " . . . increments in the frequency and intensity of responses already learned by an individual, shown in the presence of other individuals usually engaged in the same behavior" (p. 411). We shall use the term in the same manner in the review which follows.

Fish. There have been a number of recent reports substantiating earlier observations of the effect of one fish on another (**15**, **45**, **223**, **224**). There is a lower rate of locomotor activity for a group of goldfish than for an individual (**45**, **223**, **224**), but this effect appears to be dependent upon a visual response to like-formed individuals. Allee (**15**) labels this group effect "automatic cooperation" or "unconscious mutualism."

Birds. Mowrer (**165**) reports that birds being trained to talk are inhibited by the presence of another bird, although talking is facilitated by the presence of the trainer. It is likely that bird calls are facilitated by the presence of other birds in a nonlaboratory situation.

Infra-primate mammals. A number of studies on social facilitation have been done on the rat in the past decade. Rasmussen (**177**) found that a social situation influences drinking for the rat, in that rats drinking in groups of three drank more frequently that rats drinking in isolation. Bruce (**49**) obtained similar results using water as the reward for mastering an open field problem. When a second rat was present in the field, the time for the first drink was shortened whether both animals or only one had access to the water. In a later study (**50**), involving variation of the drive strength, no facilitation was found.

Using food as the reward, Winslow (**244**) found that speed of running on a straight runway, and on two identically designed elevated mazes, was affected by the presence of a second rat. The competitive situation facilitated the behavior of some subjects and inhibited the behavior of others (in some instances to the point of a refusal to perform). The rats that were continuous early winners were facilitated by the partner, while those that were continuous early losers reduced their efforts in later trials. There were no observed differences between cage and noncage mates. Winslow (**245**) has obtained similar results with cats trained to run singly, then in pairs in a 25-foot runway. Some of Mowrer's (**162**, **164**) work on competition and food sharing in the rat bears on this problem. Mowrer finds that,

> When animals which have learned individually to run a long alley for food are run in pairs, the initial effect of the resulting competitiveness is to increase

the speed of both animals slightly; but soon the slower of the two becomes demoralized (extinguished) and either stops running entirely or runs very erratically (**164**, p. 3).

Competition was changed to sharing by manipulating the animal-pellet weight ration. If the rat had an opportunity to escape with the food pellet into a compartment which excluded his cage mates, competition continued and no food sharing appeared. Lepley (**148**) failed to confirm the results of his earlier study in which rats that were paired showed a small amount of facilitation in running speed.

In all these studies of mammals the subjects have been trained to a criterion singly and then run singly and/or in pairs to determine the amount of facilitation attributable to the pairing. The amount of facilitation obtained generally has been negligible. Studies in which social facilitation is related to the learning process might reveal that this factor is of greater importance in mammalian social behavior than has been revealed by previous studies.

Using pairs of rats, one reared in isolation and the other in a group, equated for swimming ability, Bayroff (**23**) failed to find a significant difference between the isolate and group-reared subjects in speed of swimming underwater against a current. The factors of age, early experience, weight, and litter membership appeared to have no effect in this experimental situation.

A somewhat different type of problem situation has been used by Miller and Postman (**157**), Ross, Smith, and Denenberg (**186**), and Denenberg (**88**) to determine the influence of a group on the (learned?) performance of an individual. The question investigated in each of these studies was, "Will rats hoard more in groups than they will in isolation?" Using a central bin of food pellets, Miller and Postman found no significant increase in the amount hoarded to the rat's "home" cage. In the study of Ross *et al.* (**186**), rats hoarded under both solitary and group conditions over a period of eight weeks. The subjects were isolated from weaning until they were approximately four months of age. They were then run in isolation and in groups in a random manner so that each rat was paired with every other rat for part of the eight-week period. There was no facilitation of hoarding as a result of the group situation. Instead, a dominance relationship occurred in which the submissive member of the pair hoarded less under the group condition. Using a larger number of subjects, all reared under group conditions, Denenberg (**88**) found that hoarding was significantly greater under isolation conditions. This result is attributed to the social facilitation of eating under the group condition, thus reducing the amount of time devoted to pellet carrying.

Rats placed in novel situations, singly and in pairs, were not emotionally influenced by the group situation, as determined by counting and weighing the feces, according to Anderson (**18**).

Roubaud and Provost (**188**) found that sheep and goats when isolated from a group exhibited an increase in temperature and a reappearance of trypanosomes in the blood. This effect is said to be dependent on the agitation which results from the sudden change of environment. The group situation apparently has an indirect inhibiting effect on the infection.

Ross and Ross have reported observations on the social facilitation of feeding in dogs. In the first study (**183**), the amounts eaten in isolation and with the litter are compared. A litter of four Chow-Basenji puppies ate 14 per cent more in the group situation, while a litter of Terrier-Dachshund puppies ate 51 per cent more in the group situation. Individual performance varied from a 3 per cent increase to an 86 per cent increase in the food consumed under the group feeding situation. No reliable measures of dominance relationships were found.

In a second study (**184**), the effect of feeding hungry puppies on the eating behavior of satiated puppies was investigated. The results appeared to indicate that the presence of hungry puppies facilitated further eating in satiated puppies. The increase in the amount eaten by each individual varied from 30 to 200 per cent. Fellenius (**99**) has observed a facilitative effect on food preferences in dogs as a result of the presence of a gull in the feeding situation. Brogden (**47**) found no evidence of more rapid conditioning in dogs when a second dog was introduced into the laboratory.

IMITATION

Crawford (**77**) uses the term imitation to refer "to observational learning in problem situations" (p. 411). The distinction between social facilitation and imitation is somewhat difficult to maintain in certain cases. In a general review of the literature on imitation in animals, Roberts (**181**) concludes that imitation has been proved only in birds, where it is dependent on auditory cues.

Fish. In 1940, Shlaifer and Breder (**225**) reported finding a tendency of a social nature for tarpon to rise in imitation of manipulated models. In a later study, Breder (**43**) found that this imitation was less frequent in adult animals. It is doubtful that this behavior pattern in young tarpons is imitation in the sense of the definition given above.

Birds. Akhmeteli (**1**) has found that pigeons can learn food-getting behavior, without feeding, from a trained partner. The pecking movements in the trained bird appear to act as the stimulus for pecking in the untrained bird. The rate of acquisition of the behavior was slow and unstable. It is possible that this represents social facilitation without true observational learning as an important determinant.

Infra-primate mammals. Bayroff (**24**) and Bayroff and Lard (**25**) have investigated learning in the white rat in a water maze. The rats were required to select the correct path to a goal, the only cue to the correct path being a rat that had received previous training in the problem. With food as the reward, all rats eventually learned to follow a leader. In another well-known experiment, Miller and Dollard (**159**) have demonstrated that although there was no initial tendency for the rats to imitate, they did learn that the direction which a leader takes signifies a reward. When a leader with a different color was substituted, there was still a marked tendency for this leader to be followed, but it was not as strong as the initial tendency.

Brogden (**47**) investigated the effect on one dog of observing another dog undergoing conditioning of forelimb-flexion, and found no evidence that the observer could be conditioned any more rapidly because of his experience.

In a series of five experiments using thirteen cats as subjects, Herbert and Harsh (**135**) found that on problems within the subject's range of ability, the cats benefited more from observing the learning process than they did from observing trained leaders.

Primates. Imitative behavior in cebus and rhesus monkeys has been investigated using the Warden duplicate-cage technique (**240**). A total of 144 trials was given, approximately three-fourths of these resulting in imitation on the first trial. The remainder of the trials were cases of varying degrees of imitation. There is a reasonable doubt that the stimulus pattern used in this technique was well controlled. Crawford (**77**) believes that the scoring of the responses made by the subjects in the duplicate-cage technique was biased in favor of imitation.

Voronin (**238**) has reviewed the literature on the formation of conditioned responses by imitation and attempts to show that "imitation in higher nervous activity plays an essential role in the establishment of conditioned connections" (p. 131). Some data from experiments on monkeys are used to support this point.

A series of tests used by Elder (**92**) to study imitation in chimpanzees appeared to reveal that a dominant animal is a better demonstrator than one subordinate to the observing chimpanzee. White (**243**), in an essay on the use of tools by primates, proposes that primates have failed to develop a culture, not because they cannot imitate, but because they are unable to use symbols to provide continuity of experience from generation to generation.

DOMINANCE, AGGRESSION, AND TERRITORIALISM

These topics have apparently continued to interest investigators more than any other subject covered in this review. There are several excellent reviews of recent work on vertebrates, the most comprehensive of which is the one by Collias (**69**). For other general discussions and reviews, see Collias (**71**, **72**) and Allee (**10**). As it is not feasible to review every reference in detail, we shall restrict our discussion to what seem to us to be some of the more important papers. It is encouraging to note that a much wider range of species has been investigated in the last decade than in previous years. There is some difficulty involved in avoiding references to work which has been done involving the relationship between sex and dominance, since sex appears to play an important role. The best single source of material on this aspect of the problem is Beach (**27**).

Fish. Braddock (**41, 42**) has reported some of the dominance relationships existing for *Platypoecilus maculatus*. Most important of these factors is the effect of prior residence (**42**). It was found that the intruder of a pair tended to be temporarily subordinate to the resident of the aquarium even when the intruder was larger than the resident. This relationship held for both sexes. Greenberg (**115**) found that 27 groups of four sunfish each, developed relatively stable dominance hierarchies. In 17 other groups several approximately equal fish set up and defended territories, thus keeping the remainder of the group subordinate. Such factors as sex, size, and conditioned learning appear to be related to dominance for this species.

Tinbergen (**231, 232**) and Pelkwijk and Tinbergen (**175**), working on the stickleback, have found that such factors as the red belly of the male, form of the other fish, and appropriate movements serve as the cues which release aggressive responses. These researches have been based on the Lorenzian concept of releasers and hence have minimized the role of early conditioning.

Reptiles. Greenberg (**114**) has made a laboratory study of *Crotaphytus collaris*, the collared lizard, and reports that only the males are aggressive. The males showed marked discrimination of sex, probably on the basis of the throat color of the male. A ten-year study of the chameleon by Greenberg and Noble (**116, 117**) contains some interesting material bearing upon the present topic. The males defend territories during the breeding season although there may be frequent disputes over boundaries leading to temporary dominance hierarchies. Sex discrimination by both male and female is excellent, probably on the basis of distinctive coloration, structure, and movement.

Birds. For general references on dominance and other social factors in birds, see Armstrong (**21, 22**), Lack (**145, 146**), and Noble (**170**). Extensive observations have been made on common domestic fowl by Allee and his co-workers. For a detailed review of this work, see Allee (**10**). Guhl (**118, 119**) studied the relationship between peck-order position and social discrimination in small flocks of White Leghorns and found that individual differences and previous pair contact experiences were important factors in social discrimination. Guhl and Allee (**121**) rotated hens through several flocks and found that hens in unstable flocks ate less, lost weight, and laid fewer eggs. Stable dominance relationships within a flock would seem to have important survival values (**10, 123**). In a later report, Guhl (**120**) confirmed the fact that new birds added to a flock upset the flock temporarily. Douglis (**91**) found that recently added hens which were high in dominance were integrated into the flock within a few days. Hens of low dominance rank were not admitted for several weeks. Hens may be members of several flocks simultaneously and hold a different rank in each flock (**90**).

Guhl and Warren (**122**) report that dominant cocks mate more frequently and sire more chicks than their subordinates. Allee *et al.* (**12, 13, 14**) have investigated the effects of estradiol, testosterone, and thyroxin with varying results on dominance. Collias (**68**) has studied the factors determining success in initial encounters in hens, holding social facilitation, sex, and familiarity with the fighting pen constant, and reports that success is related to male hormone output, thyroxin, weight, and rank in the home flock. These four factors gave a multiple correlation coefficient of plus .75 with initial success in fighting.

Hale (**124**) observed debeaked hens and found no indication of lessened social tensions. The debeaked hens were successful in maintaining social position and in winning fights. There was a higher frequency of pecks in the debeaked hens than in normal hens.

Ringdoves have a social organization like that of pigeons, according to Bennett (**33**). The social order is purported to be dependent upon recognition factors although no permanent effects occurred from dyeing the plumage and altering the contours of the doves. Testosterone-injected males and females were found to advance in social status (**32**).

Dominance orders and territorial factors have been studied in several birds under field or near field conditions. Jenkins (**143**) has observed interspecies dominance for several groups of geese and wood ducks. Juncos and tree sparrows have been studied by Sabine (**189**) as have quail by Howard and Emlen (**130**). Other groups studied include chickadees (**127**), penguins (**46**), grouse (**216**), pheasants (**70**), Hawaiian birds (**17**), and White Pekin ducks (**11**).

Infra-primate mammals. Burt (**53**) has reviewed the literature on territoriality and discussed some of the concepts of this and the home range phenomena.

An extensive literature on dominance relationships in highly inbred strains of mice has appeared over a ten-year period, much of this work being done at the R. B. Jackson Memorial Laboratory, Bar Harbor, Maine. One of the most useful studies was done by Warne (**242**) who made a time analysis of 17 groups, each containing four male mice. The observations were made over a period of a year and a half. The results showed that only .24 per cent of the observation time was spent in actual fighting, indicating that this activity is relatively minor in the daily routine of these mice. It is interesting to note the dominant mice spent more time in grooming others than did their subordinates.

Scott (**195, 196**) has investigated genetic differences in social behavior for several strains of mice, and found indications of hereditary differences in fighting behavior. Territoriality in mice is said to be of value chiefly in controlling population density (**202**). Territories in this study were fluid and variable, not at all well defined. Scott has proposed that social behavior determines social organization (**200, 203, 205**), i.e., knowing the social behavior of given subjects, one can predict the type of social organization which will be formed.

Ginsburg and Allee (**112**) have investigated the effects of conditioning on dominance-subordination in mice. They have succeeded in demonstrating that a mouse can be conditioned to be either more or less aggressive, although it is easier to condition for less aggressiveness. Inexperienced mice can be conditioned to a low status by interfering with their adjustment to trained fighters (**207**). Scott (**208**) has found also that mice presented first with a trained fighter and then with helpless mice attack more frequently than mice presented only with the helpless mice. Uhrich (**236**) has found that previous experience was more important for albino mice than social status and weight in the outcome of a fight. Fredericson (**105**) succeeded in inhibiting fighting by stabilizing the primary reinforcing stimulus, food, by securing it in a permanent position. The effects of food deprivation (**104**), thiamin (**31**), amphetamine (**65**), and male hormone (**30**) on the fighting behavior of mice have also been investigated.

For an extensive review of the work on dominance in rats, the reader is referred to Munn (**167**), who considers in detail the work of Miller and Dollard (**159**), Miller (**158**), and Seward (**218, 219, 220, 221**). One of the interesting studies of the Norway rat in a natural habitat has been done by Calhoun (**54, 55**). A square area with 100-foot sides, surrounded by a rat-proof fence, was used in this study. Rats were tagged and marked for observation from a tower. At the time of publication, the study had been under way for eighteen months. Of importance here are these observations:

> . . . rats born near the source of food grow more rapidly, and, conversely, those born distant from the food source grow more slowly; the latter, as a result of the fact that greater weight is a deciding factor in combat, are forced to remain in submarginal areas, (with reference to the central food source) and their young in turn tend to assume a low social status. Group continuity in space and time is of value in reducing strife (**56**, p. 1122).

Collias cites an unpublished study by Heath (**132**), in which early weaned rats exhibited significantly more aggressiveness than rats that remained with the mother. There was no significant relationship between weight and aggressiveness. Rats raised in groups were found to "freeze" more frequently in problem situations than rats reared in isolation (**180**). Nearly all of the "freezers" were submissive in fighting in the cage situation. Hall and Klein (**125, 126**) have found aggressiveness in rats to be a relatively stable characteristic even when the rat is paired with another in the latter's home cage. Billingslea (**36, 37, 38**) reported on a series of experiments on the relationship between emotionality and some other behavior patterns, including aggressiveness. Related to this is Daniel's critique (**84**) of the O'Kelley-Steckle reaction. Daniel found that rats were stimulated to fight while being shocked, but little fighting occurred following shock termination. Defense postures were common in the postshock period. A similar result occurs following injections of amphetamine in mice, according to a study by Chance (**65**). In another study, Bruce (**51**) found that a rat that was dominant under food motivation was also dominant when motivated by thirst. Dominance occurred only when a competitive element was present. Fredericson (**105**) has reported similar results for mice. A film by Mowrer *et al.* (**166**) shows the development of a dominance hierarchy from the stage of no fighting, through competition, to savage fighting as hunger increases.

James has continued his study of dominance in dogs and has found that shifting dogs of known social status in one group to another group led to changes in rank only for those dogs in the middle ranks in the original group (**140**). In a later study, James (**141**) attempted to change the status of an underweight, submissive member of a litter by special handling and feeding. Even when the submissive dog equalled the dominant animals in weight, his social status remained the same. Several months in a submissive role apparently cannot be overcome by simply reducing the weight differential. This particular litter of four puppies had developed a food dominance differential by 85 days of age (**142**). This relationship was maintained until 115 days of age, when the one female became dominant over the three males. Berg (**34**) found that chronic B_1 deprivation leads to a lowering of social status. This status was regained after a series of B_1 injections.

Scott (**211**) has investigated genetic differences in the social behavior of two breeds of dogs, Basenjis and Cocker Spaniels. A standardized handling procedure was followed and all reactions were recorded at 5, 7, 9, 11, 13, and 15 weeks of age. Scott concluded that any differences between the two groups were genetically determined. In his studies of the development of social behavior in dogs, Scott has observed the development of dominance in litters of puppies (**215**). Vogel (**237**) has reported a study of social relationships, including dominance, among Springer Spaniels. The Menzels (**156**) have observed the behavior of dogs and cats of varying degrees of tameness and found that aggression ranged from mutual tolerance to constant fighting.

In studying experimental neuroses in dogs, Anderson and Parmenter (**19**) found that dogs which were in a state of anxiety had a low social status in the group.

The lair dwelling behavior of dogs has been observed by Ross (**182**). Daily defecation records were kept and these indicated that defecation tended to take place in approximately the same cage area every day. In the case of dogs that had access to an outdoor exercise cage, defecation never occurred in the living cage during the period of observation. This defecation behavior is undoubtedly related to territorial factors. Hyman (**139**) reports that hyenas in captivity have been observed to smear an anal gland secretion around the cage and pen. This is probably related to territory demarcation.

Winslow (**245**) trained cats singly to run in a runway, in a problem box, and in a string-pulling problem. When cats were run in a group of three, displaced aggression was observed. The more dominant loser commonly attacked his subordinate. An experimental study of dominance hierarchies for four groups of four cats each has been reported by Masserman and Siever (**155**). Once a hierarchy was formed, the alpha animal did not become aggressive until displaced by another animal or until a neurosis occurred as a result of motivational conflict. When drugs were administered, or the neurosis released, the animal regained his alpha status.

Dominance reactions in a flock of goats are described by Scott (**206**). The degree of aggression for each member of the flock was classified on a five-point scale ranging from 1, absence of aggressive behavior by either one of the pair, to 5, aggression by one goat with no resistance from the other. Stewart and Scott (**228**) report a study of leadership and dominance in which the criterion of dominance was the number of fighting acts, whether aggressive or submissive. Using the paired comparison method, with a restricted feeding technique, Scott (**209**) has attempted to test the frustration-aggression hypothesis on goats. The flock was composed of fourteen goats of several ages, both males and females. Frustration is said to occur when a hungry goat is prevented from eating by the presence, in a restricted feeding situation, of a more dominant subject, i.e., one who eats the limited quantity of food. The effect of delaying the feeding and recent feeding in relation to the degree of dominance was investigated. It is proposed that aggression can be controlled by controlling the amount of food given. Increasing the feeding decreases aggressiveness. Frustration (in feeding) leads to aggression only in those animals that are "in the habit of being aggressive." The relationship between dominance and control of movement was studied by Ross and Scott (**185**). It was concluded that dominance in goats is not an absolute form of social control, since it resulted in definite control of the movements of other animals in only 0.7 of the cases. The importance in dominance of such factors as age, weight, sex, and hornedness was confirmed.

Scott (**197, 198, 199, 201, 204**) has reported several investigations of the social organization of domestic sheep. The social behavior of sheep is considered in terms of mutual imitation, epimeletic behavior, signalling for care and attention, sexual behavior, and fighting. Very little fighting over food took place in the flock of sheep, although fighting among rams when a ewe was in heat was common (**204**). He says, "It appears that rewards, training and stimulation of innate behavior patterns were far more important types of control in this flock of sheep than was dominance. . . ." (p. 18). Katz (**144**) has demonstrated a dominance order in a herd of Barbary sheep. The male was dominant over both females and young. The weight and horns of the male appear to be important determinants.

That the "hook-order" of cows can be changed by removal of the horns has been reported by Woodbury (**247**). If all the horns of the herd are removed, a more peaceful social organization based on a "butt-order" develops. A dominance hierarchy among lions in an outdoor enclosure was observed by Cooper (**75**). Burt (**53**) and Gordon (**113**) have reported observations on dominance and territorialism in squirrels. The formations of herds and the importance of threatening postures and movements in zebras, horses, and mules are described by Antonius (**20**). Schenkel (**190**) reported on the importance of expressive behavior in wolves in captivity. The behavior of a muskrat population subjected to periodic drought over a fifteen-year period has been described by Errington (**96**). As the environment became more intolerable, aggressiveness increased markedly and territories broke down. Brown (**48**) reports a relationship between sexual dominance and the magnitude and stability of a conditioned response. When reversals of sexual dominance occurred, changes in the conditioned response also occurred.

Primates. Excellent studies of primate groups have been done by Carpenter (**57, 58, 60**). This investigator is also responsible for four methodological and theoretical papers on primate societies (**59, 61, 62, 63**). Several papers are available concerning the relation of sexual status to dominance (**39, 40, 66, 67, 174, 248, 249**).

Maslow's report (**154**) on infrahuman primates includes a comparative discussion of anthropoids, and Old World and New World monkeys. New World monkeys are relatively low in dominance while Old World monkeys are extremely aggressive, this being exhibited in bullying and assault. Chimpanzee dominance is reported to be playful and friendly. The behavior of paired green monkeys in problem solving is dependent almost entirely on the dominance relationship between the subjects (**93**).

A study of several social factors has been made by Warden and Galt (**241**) for nine monkeys: four cebus, three rhesus, and two sooty mangabeys. Two baited boxes with cords attached had to be pulled in by the paired subjects. A four-point dominance scale was used for rating purposes. These writers report: "The dominance factor showed itself both in the attitude of an animal to its mate and its reaction toward the food. The former might include threatening, striking, and fighting activity, and the latter an attempt to pull in both boxes and secure all the food" (p. 219). Interspecies pairings on the problem resulted in the rhesus monkeys' being dominant over the cebus in practically 100 per cent of the trials, and dominant over the mangabeys in the majority of trials. Cebus and mangabeys, when paired, resulted in higher status for the mangabeys in five out of six cases. Interspecies dominance was more stable than that within the species.

A study of the social behavior of eight female chimpanzees in a noncompetitive situation indicates that of the many types of behavior exhibited, only frequency of grooming was related to dominance in a feeding situation (**78**, **81**). Crawford (**80**) found that when chimpanzees that have been separated for varying intervals are introduced into the same cage, the animal that enters the other animal's cage is groomed first, and the one that attacks or threatens attack is likely to be dominant in a food-division test.

Montagu (**161**) reports that he has found an almost perfect relationship between body-size, waking or sleeping habits, and the origins of social life in primates. The appearance of dominance patterns appears to be related to a shift from nocturnal habits.

An extensive study of preference for a companion and dominance has been made by Nowlis (**172**) for five pre-adolescent chimpanzees. The test situation consisted of permitting a test animal to approach one of two stimulus animals and to have contact with the animal approached. Two series of tests were conducted, the first when the animals were kept in isolation at all times, except during the testing, and the second when the animals were isolated only during the day. Dominance was determined by a competitive food situation and by priority of escape from a strange situation. No relationship was found between dominance and companion preference. Nowlis states that: ". . . The behavior of the subjects suggests that aggression is the most important factor in their competitive behavior and in the development of relative dominance status. . . . The animal which consistently competes successfully by means of aggression directed towards its partners acquires habits of approach and aggression in competitive situations and furnishes conditions for the development in its partners of habits of withdrawal from punishment and from competition" (p. 52).

The relation of sexual status to dominance is discussed in a paper by Yerkes (**249**) based on a study of twenty-one pairs of adult chimpanzee mates. He says: "The basic principle of relationship revealed by the experiment involves sexual status. When a female is sexually receptive it is her *right* (italics added) to claim control of the situation without interference by her mate. . . . Sexual status, then, appears to be the most important single determiner . . . in the experiment" (p. 185).

This pattern of behavior also has been observed by Carpenter for primates other than chimpanzees (**59**). Finally, Nowlis (**173**) has studied the relationship between degree of hunger and competition in chimpanzees. The effects of satiating the dominant partner or the submissive partner is considered.

FAMILY LIFE[1]

Infra-primate mammals. Reference has been made previously to the research program of the Jackson Laboratory where part of the effort is being directed toward a long-term study of the development of the dog (**108, 187, 210, 211, 226**). By standardizing the handling technique, a relatively constant environment is provided, against which genetic differences in behavior can be compared. Royce's (**187**) factorial study of several physiological, psychological, and social measures, and Scott and Marston's (**214**) discussion of critical periods of development are both related indirectly to the family life of the dog. Extensive observation of several litters of dogs has led Scott and Marston to propose that those periods in the development of an organism in which new social relationships are formed are "critical" in the sense that the habits formed in the critical periods are likely to be carried over into later life. The experimental program testing this hypothesis is just now starting. The only other study in this group is by Seward (**217**), who has pointed out the importance of hunger in the filial behavior of guinea pigs.

Primates. Carpenter (**59**) has summarized his observations on grouping patterns and family relationships for several nonhuman primate societies. This writer's observations on play and its relationship to aggressiveness also are of importance. Beach (**26**) has reviewed some of the current concepts of play. (See Shadle, **222**, also.) Spence (**227**) reports that five out of seven mother chimpanzees were able to recognize their own children when the child was presented along with an unrelated offspring following varying periods of isolation.

EFFECT OF ISOLATION ON SOCIAL BEHAVIOR

Infra-primate mammals. The studies of Denenberg (**88**) and Ross *et al.* (**186**) on group hoarding indicate that rats reared in isolation and hoarding in groups were more overtly aggressive than rats reared in groups and hoarding in groups. Heath's (**132**) study of rats isolated early from the mother is also pertinent to this point. Scott (**204**) notes that two lambs isolated from the flock and bottle-fed were not accepted by the mothers upon return to the flock.

Primates. Reference has already been made to Crawford's (**80**) study of pairs of chimpanzees isolated from each other for from five weeks to three years, and to Nowlis's (**172**) study of companionship preference which inolved temporary isolation from the group.

COMMUNICATION

A review of the types of communication that may exist among animals and between man and animals has been presented by Révész

---

[1] A stimulating discussion of the development of family life in infrahuman animals has been given by Schneirla (**194**).

(**179**). Mowrer's (**165**) chapter on the psychology of "talking birds" is a particularly provocative presentation of the technique of teaching birds to talk and the relationship of this behavior pattern to the development of language in the human. It is proposed that the birds reproduce those sounds which "sound good to them" (p. 699). The acquisition of words is considered in terms of primary and secondary reinforcement. Hayes (**131**) has reported in abstract on her work in training a chimpanzee to speak. Allard (**2**) has made field observations on starlings and mockingbirds and has found an unusual range of sounds which normally occur in the summer environment being reproduced "out of season," as it were.

Carpenter (**57**) recorded gibbon calls in the field and then played some of the recordings back. Vocalization is said to be an important part of the behavior of this group. Although no detailed analysis of gibbon calls has been reported, Carpenter does note that, "In general it seems that there are five types or series of sound patterns. Each series is distinct, is produced in a specific situation and produces in associated animals characteristic responses. Several times it was possible to stimulate the wild animals to call by playing back to them the recordings of their own calls which had just been made" (p. 325).

RECOGNITION OF KIND

Several references have been made in this review to factors which appear to be determinants of the ability of some fish, lizards, hens, mammals, and infrahuman primates to recognize other members of a group. This recognition is most frequently on the basis of distinctive coloration, form, movement, or vocalization. In an infrahuman primate group there appears, according to Carpenter, to be a recognition by every member of every other member of the group, although this has not been experimentally demonstrated.

Howells (**136**) and Howells and Vine (**137**) in a study of social learning have contributed some data on this problem. Incubator-hatched bantams and leghorns were reared in mixed groups, thus giving each member of the group opportunity for equal experience with bantams and leghorns. When the subjects were three weeks of age, they were trained in a Y discrimination box with certain members of its flock as cues. The chicks learned to go to their own kind more readily than to the non-like bird. This effect is interpreted as "a product of a differential influence on the learning process exerted by the innate factors" (**137**, p. 547). Bantams learned to go to the bantams more readily than leghorns learned to go to their own kind, perhaps because of greater activity in the bantams.

COOPERATION AND FOOD SHARING

Infra-primate mammals. Daniel (**82**, **83**, **85**, **86**) has conducted several studies on cooperation in rats. The problem was arranged so that a pair of rats could escape an electric shock by cooperating in a feeding situation (**83**). The individual animals were trained in the problem so that it was essentially a case of rats learning to cooperate. As the rats worked in the problem, they tended to pay more attention to each other so that exchange of position from food to platform was accomplished frequently enough to provide satisfaction of both the motives of hunger and escape from shock. An extension of this study (**85**) indicated that cooperation could be maintained with low shock or threat of shock but would disappear with no shock at all.

In a well-known film, Mowrer (**163**) has shown that three rats faced with a problem of manipulating a bar at one point in order to obtain food at a bin at another point established a "parasitic" relationship rather than a cooperative one. Tsai (**235**) reports that rats learned to cooperate in a problem requiring the simultaneous pressing of separate keys in order to raise a door to secure food. Rats also learned to press a key to release a partner from a detention box before proceeding with the remainder of the problem. In a problem requiring rats to pull a cord together in order to obtain food, Winslow (**244**) found no clear evidence of cooperation.

Mowrer (**162**, **164**) reports that in one experiment feeding behavior in rats passed through several stages, including food-sharing. If deprivation is increased, competition becomes more intense and sharing does not develop.

Primates. Monkeys failed to exhibit spontaneous cooperation in a problem requiring the pulling of a weighted box too heavy for either member of the pair to move singly (**241**). The animals failed to cooperate following tuition, contrary to results obtained for chimpanzees. Wolfle and Wolfle (**246**) have studied cooperative behavior genetically in a comparison of monkeys and children. Crawford (**79**) studied the ability of six young chimpanzees to cooperate in solving a serial response problem. Varying degrees of cooperation were shown.

Carpenter and Locke (**56**) report the development of trading responses in a cebus monkey which learned to exchange a specific object for food. This behavior appears to be related to a type of sharing. Nowlis (**172**) found that food transfer occurred from a subordinate to a dominant animal. There was no relationship between the sharing of food and preference for a companion. Yerkes (**250**) reports that food was shared by partners only when the female was in estrus and took the food as her right.

GROOMING

Warden and Galt (**241**) report that unilateral brain injury decreased grooming in monkeys and bilateral injury completely eliminated grooming. Grooming often took precedence over food seeking and did not seem to be related to sex or dominance status. Crawford (**80**) found that one index of dominance status was grooming, the grooming being done by the subordinate animal. In young chimpanzees " . . . 87 per cent of the social behavior occurred in the form of two distinct patterns, grooming and play-fighting" (**172**, p. 33).

Reference should be made to several other studies related more or less directly to the social behavior of vertebrates. Attention might be called to Hebb's (**133**, pp. 245–250) description of mental illness in chimpanzees since this appears to be social in nature. Some of the work on neurotic behavior (**110**, **153**, **156**) has important implications in social development. Hebb's (**134**) proposed methods of analysis of temperament are of interest. The effect of a conflict situation on pairs of chimpanzees is described in an abstract by Fletcher (**101**).

Finally, we can include here Bechterev's (**29**) report on the "influence" of man on the behavior of animals, and Curti's (**74**) report on the responses of albino rats to the presence of a cat. Visual cues are apparently not necessary in order for the rat to recognize a cat.

SUMMARY AND INTERPRETATION

Numerous papers in recent years have presented various theoretical interpretations of social behavior. The most popular "explanatory" principles have been those of cooperation and competition (**3**, **5**, **6**, **7**, **8**, **9**, **10**, **94**, **95**, **109**, **160**), both of which have sometimes been used as intervening variables and at other times as dependent variables. There have been theories based on releasers (Tinbergen, **229**, **230**, **231**, **232**, **233**) (Lorenz, **149**, **150**, **151**, **152**), trophallaxis (**191**, **192**), and others in which learning (**76**, **137**, **138**, **158**, **159**) is the primary principle proposed as a basis for a theoretical comparative social psychology. Each of these theories has some merit, but all suffer from too heavy an emphasis on a narrow range of methods and problems for which their theory has some "face" validity.

Beach (**28**) has pointed out the need for a comparative study of social behavior and for the advantage of using a wide variety of species as test subjects. It is difficult to see how there can be a comparative social psychology unless and until a wider range of animal subjects is studied. But the use of more species will be of relatively little value unless the investigations undertaken are guided by some systematic theoretical framework such as have been proposed for learning by Hull and for personality by Sears. A framework of this kind might entail the formulation of a number of postulates and the systematic investigation of the intervening variables which are deduced. Unifying principles are urgently required if such studies are to have a major applicability to the problems of human social behavior.

In addition to deficiencies in theoretical orientation and in the range of species studied, there also exist weaknesses in the techniques used in the study of social behavior. There is a need for new approaches to existing problems, and for new problems to be conceived and studied. The use of models by those individuals investigating the concept of social releasers represents an attempt to meet such needs. It is very possible that much of social behavior is situational; hence the range of experimental situations used might well be increased far beyond runways, mazes, and problem boxes. One of the most promising attempts at the solution of such problems is to be found in the symposium on the "Methodology and Techniques for the Study of Animal Societies" (**212**). This report documents an attempt to approach the study of social behavior in a systematic manner by delineating problem areas and by suggesting appropriate techniques to be used.

Finally, the range of problems studied has been relatively limited, as can be seen easily by examining the various areas for which work is cited in the review. Little has been done on such important phenomena as play, communication, imitation, and learning in social situations. These phenomena are of great interest and perhaps the next decade of research will provide more significant advances in our knowledge.

BIBLIOGRAPHY

1. AKHMETELI, M. Imitation in pigeons. *Trans. Berritashvil Physiol. Inst.*, 1941, **4**, 345–365. (*Psychol. Abstr.*, 1943, **17**, No. 3024.)
2. ALLARD, H. A. Vocal mimicry of the starling and the mocking bird. *Science*, 1939, **90**, 370–371.
3. ALLEE, W. C. *The social life of animals.* New York: Norton, 1938.
4. ALLEE, W. C. *Djurens sociala liv.* Stockholm: Natur ach Kultur, 1940. (*Psychol. Abstr.*, 1941, **15**, No. 1401.)
5. ALLEE, W. C. Concerning the origin of sociality in animals. *Scientia, Milan*, 1940, **67**, 154–160.
6. ALLEE, W. C. Group organization among vertebrates. *Science*, 1942, **95**, 289–293.
7. ALLEE, W. C. Human conflict and cooperation; The biological background. In L. Bryson, L. Finklestein and R. M. MacIver (Eds.), *Approaches to national unity.* Fifth Conference on Science, Philosophy, and Religion, 1945, 312–364.
8. ALLEE, W. C. Animal sociology. *Encycl. Britannica*, 1947.
9. ALLEE, W. C. Extrapolation in comparative sociology. *Scientia, Milan* 1949, **6**, 135–142.
10. ALLEE, W. C. Dominance and hierarchy in societies of vertebrates. Univ. of Chicago. Unpublished Manuscript, 1950.
11. ALLEE, W. C., ALLEE, M. N., RITCHEY, F., & CASTLE, E. W. Leadership in a flock of White Pekin ducks. *Ecology*, 1947, **28**, 310–315.
12. ALLEE, W. C., & COLLIAS, N. The influence of estradiol on the social organization of flocks of hens. *Endocrinology*, 1940, **27**, 87–94.

13. Allee, W. C., Collias, N. E., & Beeman, E. The effect of thyroxin on the social order in flocks of hens. *Endocrinology*, 1940, **27**, 827–835.

14. Allee, W. C., Collias, N. E., & Lutherman, C. Z. Modification of the social order in flocks of hens by the injection of testosterone propionate. *Physiol. Zool.*, 1939, **12**, 412–440.

15. Allee, W. C., Finkel, A. J., & Haskins, W. H. The growth of goldfish in homotypically conditioned water; a population study in mass physiology. *J. exp. Zool.*, 1940, **84**, 417–443.

16. Allee, W. C., Emerson, A. E., Park, O., Park, T., Schmidt, K. P. *Principles of animal ecology*. Philadelphia: Saunders, 1949.

17. Amadon, Dean. Ecology and the evolution of some Hawaiian birds. *Evolution (N. Y.)*, 1947, **1**, 63–68.

18. Anderson, E. E. The effect of the presence of a second animal upon the emotional behavior of the male albino rat. *J. soc. Psychol.*, 1939, **10**, 265–268.

19. Anderson, O. D., & Parmenter, R. A. A long-term study of the experimental neurosis in the sheep and dog. *Psychosom. Med. Monogr.*, 1941, **2**, 1–150.

20. Antonius, O. Über Herdenbildung und Paarungseigentümlichkeiten der Einhufer. *Z. Tierpsychol.*, 1937, **1**, 259–289. (*Psychol. Abstr.*, 1938, **12**, No. 3612.)

21. Armstrong, E. A. *Bird display; An introduction to the study of bird psychology*. Cambridge: Macmillan, Cambridge Univer. Press, 1942.

22. Armstrong, E. A. *Bird display and behavior*. London: Lindsay Drummond, Ltd., 1947.

23. Bayroff, A. G. The experimental social behavior of animals. II. The effects of early isolation of white rats on their competition in swimming. *J. comp. Psychol.*, 1940, **29**, 293–306.

24. Bayroff, A. G. A preliminary study of imitational learning in white rats. *J. Elisha Mitchell sci. Soc.*, 1941, **57**, 213.

25. Bayroff, A. G., & Lard, K. E. Experimental social behavior of animals. III. Imitational learning of white rats. *J. comp. Psychol.*, 1944, **37**, 165–171.

26. Beach, F. A. Current concepts of play in animals. *Amer. Nat.*, 1945, **79**, 523–541.

27. Beach, F. A. *Hormones and behavior*. New York: Hoeber, 1948.

28. Beach, F. A. The snark was a boojum. *Amer. Psychologist*, 1950, **5**, 115–124.

29. Bechterev, W. "Direct influence" of a person upon the behavior of animals. *J. Parapsychol.*, 1949, **13**, 166–176.

30. Beeman, E. A. The effect of male hormones on aggressive behavior in mice. *Physiol. Zool.*, 1947, **20**, 373–405.

31. Beeman, E. A., & Allee, W. C. Some effects of thiamin on the winning of social contacts in mice. *Physiol. Zool.*, 1945, **18**, 195–221.

32. Bennett, M. A. The social hierarchy in ring doves. II. The effect of treatment with testerone propionate. *Ecology*, 1940, **21**, 148–165.

33. Bennett, M. A. The social hierarchy in ring doves. *Ecology*, 1939, **20**, 337–357.

34. Berg, I. A. Comment on chronic vitamin B_1 deprivation. *Science*, 1947, **105**, 252.

35. Berwein, M. Observations and experiments on the social life of minnows. *Z. vergl. Physiol.*, 1941, **28**, 402–420. (*Psychol. Abstr.*, 1943, **17**, No. 2777.)

36. Billingslea, F. Y. Analysis of intercorrelations between emotionality and other behavior salients in the rat. *Psychol. Bull.*, 1940, **37**, 430. (Abstract)

37. Billingslea, F. Y. The relationship between emotionality and various

other salients of behavior in the rat. *J. comp. Psychol.*, 1941, **31**, 69–77.

38. BILLINGSLEA, F. Y. Intercorrelational analysis of certain behavior salients in the rat. *J. comp. Psychol.*, 1942, **34**, 204–211.
39. BIRCH, H. G., & CLARK, G. Hormonal modification of social behavior; II. The effects of sex-hormone administration on the social dominance status of the female-castrate chimpanzee. *Psychosom. Med.*, 1946, **8**, 320–331.
40. BIRCH, H. G., & CLARK, G. Hormonal modification of social behavior; IV. The mechanism of estrogen-induced dominance in chimpanzees. *J. comp. physiol. Psychol.*, 1950, **43**, 181–194.
41. BRADDOCK, J. C. Some aspects of the dominance-subordination relationship in *Platypoecilus maculatus*. *Anat. Rec.*, 1942, **84**, 508–509.
42. BRADDOCK, J. C. The effect of prior residence upon dominance in the fish *Platypoecilus maculatus*. *Physiol. Zool.*, 1949, **22**, 161–169.
43. BREDER, C. M., JR. Social and respiratory behavior of large tarpon. *Zoölogica (N. Y.)*, 1942, **27**, 1–4.
44. BREDER, C. M., JR. Innate and acquired behavior affecting the aggregation of fishes. *Physiol. Zool.*, 1946, **19**, 154–190.
45. BREDER, C. M., JR., & NIGRELLI, R. The significance of differential locomotor activity as an index to the mass physiology of fishes. *Zoologica (N. Y.)*, 1938, **23**, 1–29.
46. BRIAN, R. The breeding behavior of penguins with special reference to *Pygascelis papua* (Forster). *Brit. Mus. Nat. Hist. Sci. Rep.*, 1940, Part I, 195–254. (*Psychol. Abstr.*, 1941, **15**, No. 5241.)
47. BROGDEN, W. J. Imitation and social facilitation in the social conditioning of forelimb-flexion in dogs. *Amer. J. Psychol.*, 1942, **55**, 77–83.
48. BROWN, R. H. Stability of conditioning and sexual dominance in the rabbit. *Science*, 1937, **86**, 520.
49. BRUCE, R. H. An experimental analysis of the social factors affecting the performance of white rats. I. Performance in learning a simple field situation. *J. comp. Psychol.*, 1941, **31**, 363–377.
50. BRUCE, R. H. An experimental analysis of social factors affecting the performance of white rats. II. Effects of varying the drive. *J. comp. Psychol.*, 1941, **31**, 379–394.
51. BRUCE, R. H. An experimental analysis of social factors affecting the performance of white rats. III. Dominance and cooperation motivated by water and food deprivation. *J. comp. Psychol.*, 1941, **31**, 395–412.
52. BRUCE, R. H. A summary of four years' work on the social behavior of albino rats. *Psychol. Bull.*, 1939, **36**, 656. (Abstract)
53. BURT, W. H. Territoriality and home range concepts as applied to mammals. *J. Mammal.*, 1943, **24**, 346–352.
54. CALHOUN, J. B. The development and role of social status among wild Norway rats. *Anat. Rec.* (Suppl.), 1948, **101**, 44.
55. CALHOUN, J. B. The study of wild animals under controlled conditions. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1103–1112.
56. CARPENTER, C. R., & LOCKE, N. M. Notes on symbolic behavior in a cebus monkey. *J. genet. Psychol.*, 1937, **51**, 267–278.
57. CARPENTER, C. R. Behavior and social relations of free-ranging primates. *Sci. Mon., N. Y.*, 1939, **48**, 319–325.
58. CARPENTER, C. R. A field study in Siam of the behavior and social relations of the gibbon (*Hylobates lar*). *Comp. Psychol. Monogr.*, 1940, **16**, No. 5.
59. CARPENTER, C. R. Societies of mon-

keys and apes. *Biological Symposia*, 1942, **8**, 177–204.

60. Carpenter, C. R. Sexual behavior of free ranging rhesus monkeys. I. Specimens, procedures and behavioral characteristics of estrus. II. Periodicity of estrus, homosexual, autoerotic and non-conformist behavior. *J. comp. Psychol.* 1942, **33**, 113–142.
61. Carpenter, C. R. Characteristics of social behavior in non-human primates. *Trans. N. Y. Acad. Sci.*, 1942, **4**, 248–258.
62. Carpenter, C. R. Concepts and problems of primate sociometry. *Sociometry*, 1945, **8**, 56–61.
63. Carpenter, C. R. General plans and methodology for field studies of the naturalistic behavior of animals. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1003–1006.
64. Cattell, J. (Ed.) *Biological Symposia.* Vol. II. *Speciation, defense mechanisms in plants and animals, biological basis of social problems and regeneration.* Lancaster, Pa.: Jacques Cattell Press, 1941.
65. Chance, M. R. A. A peculiar form of social behavior induced in mice by amphetamine. *Behaviour*, 1947, **1**, 64–70.
66. Clark, G., & Birch, H. G. Hormonal modifications of social behavior. I. The effect of sex-hormone administration on the social behavior of a male-castrate chimpanzee. *Psychosom. Med.*, 1945, **7**, 321–329.
67. Clark, G., & Birch, H. G. Hormonal modifications of social behavior. III. The effects of stilbesterol therapy on social dominance in the female-castrate chimpanzee. *Bull. Canad. psychol. Ass.*, 1946, **6**, 15–18.
68. Collias, N. E. Statistical analysis of factors which make for success on initial encounters between hens. *Amer. Nat.*, 1943, **77**, 519–538.
69. Collias, N. E. Aggressive behavior among vertebrate animals. *Physiol. Zool.*, 1944, **17**, 83–123.
70. Collias, N. E. Social life and the individual among vertebrate animals. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1074–1092.
71. Collias, N. E. Some variations in grouping and dominance patterns among birds and mammals. *Zoologica*, 1950, **35**, 97–119.
72. Collias, N. E. Problems and principles of animal sociology. In C. P. Stone (Ed.), *Comparative psychology.* (3rd Ed.) New York: Prentice-Hall, 1951. Pp. 388–422.
73. Collias, N. E., & Tabor, R. D. Grouping and dominance relations among wild Ring-necked Pheasants. *Anat. Rec.* (Suppl.), 1948, **101**, 44–45.
74. Curti, M. W. A further report on fear responses of white rats in the presence of cats. *J. comp. Psychol.*, 1942, **34**, 51–53.
75. Cooper, J. B. An exploratory study on African lions. *Comp. Psychol. Monogr.*, 1942, **17**, 1–48.
76. Cooper, J. B. Learning and social behavior: A point of view. *J. soc. Psychol.*, 1950, **32**, 31–45.
77. Crawford, M. P. The social psychology of the vertebrates. *Psychol. Bull.*, 1939, **36**, 407–446.
78. Crawford, M. P. Relations between dominance and noncompetitive behavior in female chimpanzees. *Psychol. Bull.*, 1941, **38**, 697. (Abstract)
79. Crawford, M. P. The cooperative solving by chimpanzees of problems requiring serial responses to color. *J. soc. Psychol.*, 1941, **13**, 259–280.
80. Crawford, M. P. Dominance and the behavior of pairs of female chimpanzees when they meet after varying intervals of separation. *J. comp. Psychol.*, 1942, **33**, 259–265.
81. Crawford, M. P. The dominance

and social behavior, for chimpanzees, in a non-competitive situation. *J. comp. Psychol.*, 1942, **33**, 267–277.

82. Daniel, W. J. Cooperative behavior in rats. *Psychol. Bull.*, 1941, **38**, 564. (Abstract)
83. Daniel, W. J. Cooperative problem solving in rats. *J. comp. Psychol.*, 1942, **34**, 361–368.
84. Daniel, W. J. An experimental note on the O'Kelley-Steckle reaction. *J. comp. Psychol.*, 1943, **35**, 267–268.
85. Daniel, W. J. Higher order cooperative problem solving in rats. *J. comp. Psychol.*, 1943, **35**, 297–305.
86. Daniel, W. J. How cooperatively do individual rats solve a problem in a social situation? *Amer. Psychologist*, 1947, **2**, 316. (Abstract)
87. Darling, F. F. *A naturalist on Rona; Essays of a biologist in isolation.* Toronto: Oxford, 1939.
88. Denenberg, Victor. A further study of the effects of a group situation upon hoarding behavior of the white rat. Unpublished master's thesis, Purdue Univer., 1951.
89. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology*, 1945, **26**, 297–302.
90. Douglas, M. B. Hens that are members of as many as four flocks may maintain a different social status in each. *Anat. Rec.*, 1944, **89**, 23.
91. Douglis, M. D. Social factors influencing the hierarchies of small flocks of the domestic hens: interactions between resident and part-time members of the organized flocks. *Physiol. Zool.*, 1948, **21**, 147–182.
92. Elder, J. H. Tests of imitation in chimpanzees. *Psychol. Bull.*, 1939, **36**, 595. (Abstract)
93. Elder, J. H., & Nordbal, N. Behavior of green monkeys in a frustrating social situation. *Psychol. Bull.*, 1940, **37**, 581. (Abstract)
94. Emerson, A. E. Biological sociology. *Denison Univer. Bull., J. sci. Lab.*, 1941, **36**, 146–155.
95. Emerson, A. E. The biological basis of social cooperation. *Trans. Ill. Acad. Sci.*, 1946, **39**, 9–14.
96. Errington, P. L. Reaction of muskrat population to drought. *Ecology*, 1939, **20**, 168–186.
97. Essen, J. V. Die Verhaltensforderung als Grundlage gegenseitiger Wertung. *Arch. ges. Psychol.*, 1938, **100**, 485–500. (*Psychol. Abstr.*, 1939, **13**, No. 2058.)
98. Faure, A. Animal societies and human societies. *Schweiz. Z. Psychol. Anivend.*, 1947, **76**, 113–123. (*Psychol. Abstr.*, 1947, **21**, No. 4512.)
99. Fellenius, V. Observations de psychologie sociale des animaux dans leur comportment a l'égard de la nourriture. *J. Psychol. norm. path.*, 1938, **35**, 624–627. (*Psychol. Abstr.*, 1940, **14**, No. 984.)
100. Fisher, K. C. The selected temperature of Atlantic salmon and speckled trout and the effect of temperature on the response to an electrical stimulus. *Physiol. Zool.*, 1950, **23**, 27–34.
101. Fletcher, F. M. Certain behavior patterns associated with a conflict situation in young chimpanzees. *Psychol. Bull.*, 1940, **37**, 488. (Abstract)
102. Folger, H. T. Reaction of tadpoles to light. I. Effects of chemicals on aggregation of tadpoles in unilateral illumination. *Physiol. Zool.*, 1939, **12**, 136–150.
103. Fraenkel, G. S., & Gunn, D. L. *The orientation of animals: Kineses, taxes and compass reactions.* New York: Oxford Univer. Press, 1940. (*Psychol. Abstr.*, 1941, **15**, No. 1682.)
104. Fredericson, E. The effects of food deprivation upon competitive and

spontaneous combat in C57 black mice. *J. Psychol.*, 1950, **29**, 89–101.

105. FREDERICSON, E. Inhibition of competitive fighting by stabilization of the primary reinforcing stimulus. *Amer. Psychologist*, 1950, **5**, 262. (Abstract)
106. FREISLING, J. Studien zur Biologie der Wechsselkröte (Bufo virdis Laur.) *Oesterrlich. Zool. Z.*, 1948, **1**, 383–440. (*Biol. Abstr.*, 1949, **23**, No. 28260.)
107. FUHR, I., & SEYMOUR, D. S. Simulated burrow systems for studies with rodent pests. *Wildlife Management*, 1947, **11**, 150–153.
108. FULLER, J. L. Individual differences in the reactivity of dogs. *J. comp. physiol. Psychol.*, 1948, **41**, 339–347.
109. GALT, W. The principle of cooperation in behavior. *Quart. Rev. Biol.*, 1940, **15**, 401–410.
110. GANTT, W. H. *Experimental basis for neurotic behavior.* New York: Hoeber, 1944.
111. GINSBURG, B. Genetics and social behavior—A theoretical synthesis. 20th Commemoration Lectures. Bar Harbor: R. B. Jackson Memorial Laboratory, 1949.
112. GINSBURG, B., & ALLEE, W. C. Some effects of conditioning on social dominance and subordination in inbred strains of mice. *Physiol. Zool.*, 1942, **15**, 485–506.
113. GORDON, K. Territorial behavior and social dominance among *Sciurdae*. *J. Mammal.*, 1940, **17**, 171–172.
114. GREENBERG, B. Notes on the social behavior of the collared lizard. *Copeia*, 1945, **4**, 225–230.
115. GREENBERG, B. Some relations between territory, social hierarchy, and leadership in the green sunfish (*Lepomis cyanellus*). *Physiol. Zool.*, 1947, **20**, 267–299.
116. GREENBERG, B., & NOBLE, G. K. Dominance, social order and territory in the lizard (*Anolis carolinensis*). *Anat. Rec.*, 1942, **84**, 508–509.
117. GREENBERG, B., & NOBLE, G. K. Social behavior of the American chameleon (*Anolis carolinensis*). *Physiol. Zool.*, 1944, **17**, 392–439.
118. GUHL, A. M. Social discrimination in small flocks of white leghorns. *Anat. Rec.*, 1941, **81**, 114.
119. GUHL, A. M. Social discrimination in small flocks of the common domestic fowl. *J. comp. Psychol.*, 1942, **34**, 127–148.
120. GUHL, A. M. Some observations and experiments on social behavior in the domestic fowl. *Trans. Kansas Acad. Sci.*, 1945, **47**, 337–341.
121. GUHL, A. M. Social dominance and receptivity in the domestic fowl. *Physiol. Zool.*, 1950, **23**, 361–366.
122. GUHL, A. M., & ALLEE, W. C. Some measurable effects of social organization in flocks of hens. *Physiol. Zool.*, 1944, **17**, 320–347.
123. GUHL, A. M., & WARREN, D. C. Number of offspring sired by cockerels related to social dominance in chickens. *Poultry Sci.*, 1946, **25**, 460–472.
124. HALE, E. B. Observations on the social behavior of hens following debeaking. *Poultry Sci.*, 1948, **27**, 591–592.
125. HALL, C. S. Measuring individual differences in aggressiveness in rats. *Psychol. Bull.*, 1941, **38**, 706. (Abstract)
126. HALL, C. S., & KLEIN, S. J. Individual differences in aggressiveness in rats. *J. comp. Psychol.*, 1942, **33**, 371–383.
127. HAMERSTRAM, F. Dominance in winter flocks of chickadees. *Wilson Bull.*, 1942, **54**, 32–42.
128. HARDING, C. F., III. Objective studies of the social behavior of animals. *Appl. Anthrop.*, 1943, **2**, 21–39.
129. HASKELL, E. F. A natural classification of societies. *Trans. N. Y. Acad. Sci.*, 1947, **9**, 186–196.

130. Howard, W. E., & Emlen, J. T., Jr. Intercovy social relationships in the valley quail. *Wilson Bull.*, 1942, **54**, 162–170.

131. Hayes, K. J. Vocalization and speech in chimpanzees. *Amer. Psychologist*, 1950, **5**, 275. (Abstract)

132. Heath, D. H. An experiment on early weaning and aggressiveness in rats. Unpublished, 1947. (Cited by Collias, **71**.)

133. Hebb, D. O. *The organization of behaviour: A neuro-psychological theory.* New York: Wiley, 1949.

134. Hebb, D. O. Temperament in chimpanzees: I. Method of analysis. *J. comp. physiol. Psychol.*, 1949, **42**, 192–206.

135. Herbert, M. J., & Harsh, C. M. Observational learning by cats. *J. comp. Psychol.*, 1944, **37**, 81–95.

136. Howells, T. H. The innate differential in the social learning of fowls. *Psychol. Bull.*, 1939, **36**, 518. (Abstract)

137. Howells, T. H., and Vine, D. C. The innate differential in social learning. *J. abnorm. soc. Psychol.*, 1940, **35**, 537–548.

138. Hull, C. L. A primary social science law. *Sci. Mon., N. Y.*, October, 1950, **74**, 221–228.

139. Hyman, L. H. Markierungsverhalten bei der Hyäne. *Z. Morph. u. Ökol. Tiere.*, 1939, **35**, 10–13. (*Biol. Abstr.*, 1939, **13**, No. 10769.)

140. James, W. T. Further experiments in social behavior among dogs. *J. genet. Psychol.*, 1939, **54**, 151–164.

141. James, W. T. The development of dominance—Submissive hierarchy among puppies with attempts to change the position of the submissive animal. *Amer. Psychologist*, 1948, **3**, 240. (Abstract)

142. James, W. T. Dominant and submissive behavior in puppies as indicated by food intake. *J. genet. Psychol.*, 1949, **75**, 33–43.

143. Jenkins, D. Territory as a result of despotism and social organization as shown by geese. *Auk*, 1944, **61**, 30–47.

144. Katz, I. Behavioral interaction in a herd of Barbary Sheep (*Ammotragus lervia*). *Zoologica*, 1949, **34**, 9–18.

145. Lack, D. The behavior of the robin. *Proc. zool. Soc. Lond.*, 1939, Ser. A., **109**, 169–219. (*Psychol. Abstr.*, 1941, **15**, No. 198.)

146. Lack, D. *The life of the robin.* London: H. F. and G. Witherby, Ltd., 1943.

147. Langlois, T. H. A study of small-mouth bass, *Micropterus dolomieu* (Lacepede), in rearing ponds in Ohio. *Biol. Surv. Bull.*, 1936, **6**, 33. Ohio State Univer. Studies. (Cited by Haskell, **30**.)

148. Lepley, W. M. The social facilitation of locomotor behavior in the albino rat. *J. exp. Psychol.*, 1939, **24**, 106–109.

149. Lorenz, K. Der Kumpan in der Umwelt des Vogels. *J. Ornithol.*, 1935, **83**, 137–214.

150. Lorenz, K. The companion in the bird's world. *Auk*, 1937, **54**, 245–273.

151. Lorenz, K. Vergleichende Verhandlungsforschung. *Zool. Anzeiger Suppl.* (*Ver. deutsche zool. Ges.*), 1939, **12**, 69–102.

152. Lorenz, K. Vergleichende Bewegungstudien in Anatinen. *Ornithologist*, 1941, **89**, 194–294.

153. Maier, N. R. F. Experimentally induced abnormal behavior. *Sci. Mon. N. Y.*, 1948, **67**, 210–216.

154. Maslow, A. H. Dominance quality and social behavior in infra-human primates. *J. soc. Psychol.*, 1940, **11**, 313–324.

155. Masserman, J. H., & Siever, P. W. Dominance, neurosis, and aggression: An experimental study. *Psychosom. Med.*, 1944, **6**, 7–16.

156. Menzel, R., & Menzel, Rudolfine

A preliminary communication on the types of relationships between dogs and cats. *Behaviour*, 1948, **1**, 226–236.

157. Miller, G. A., & Postman, L. Individual and group hoarding in rats. *Amer. J. Psychol.*, 1946, **59**, 652–668.

158. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

159. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

160. Montagu, M. F. A. Social instincts. *Sci. Amer.*, 1950, **182**, No. 4.

161. Montagu, M. F. A. Relation between body-size, waking activity, size of eyeballs, and the origin of social life in primates. *Nature, Lond.*, 1943, **152**, 573–574.

162. Mowrer, O. H. Animal studies in the social modification of organically motivated behavior (film). State Coll., Pa.: Psychol. Cinema Reg., 1939.

163. Mowrer, O. H. An experimentally produced "social problem" in rats (film). State Coll., Pa.: Psychol. Cinema Reg., 1939.

164. Mowrer, O. H. Animal studies in the genesis of personality. *Trans. N. Y. Acad. Sci.*, 1940, **3**, 8–11.

165. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

166. Mowrer, O. H., Karnreich, J. S., & Yoffe, I. Competition and dominance in rats (film). State Coll., Pa.: Psychol. Cinema Reg., 1940.

167. Munn, N. L. *Handbook of psychological research on the rat.* New York: Houghton-Mifflin, 1950.

168. Nissen, H. W. Social behavior in primates. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 347–386.

169. Nissen, H. W. Social behavior in primates. In C. P. Stone (Ed.), *Comparative psychology.* (3rd Ed.) New York: Prentice-Hall, 1951. Pp. 423–457.

170. Noble, G. K. Symposium on the individual vs. the species. IV. The role of dominance in the social life of birds. *Auk*, 1939, **56**, 263–273.

171. Noble, G. K., & Curtis, B. The social behavior of the jewel fish, *Hemichromis bimaculatus. Bull. Amer. Mus. Nat. Hist.*, 1939, **76**, 1–76.

172. Nowlis, V. Companionship preference and dominance in the social interaction of young chimpanzees. *Comp. Psychol. Monogr.*, 1941, **17**, No. 1, 1–57.

173. Nowlis, V. The relation of degree of hunger to competitive interaction in chimpanzees. *J. comp. Psychol.*, 1941, **32**, 91–115.

174. Nowlis, V. Sexual status and degree of hunger in chimpanzee competitive interaction. *J. comp. Psychol.*, 1942, **34**, 185–194.

175. Pelkwijk, J. J. Ter, & Tinbergen, N. Eine reizbiologische Analyse einer Verhaltensweisen von *Gasterosteus aculeatus. Z. Tierpsychol.*, 1937, **1**, 193–201.

176. Rabaud, E. *Phenomene social et sociétés animales.* Paris: Felix Alcan, 1937.

177. Rasmussen, E. W. Social facilitation in albino rats. *Acta Psychol. (Hague)*, 1939, 275–294. (*Psychol. Abstr.*, 1939, **13**, No. 5315.)

178. Redfield, R. (Ed.) *Biological symposia.* VIII. *Levels of integration in biological and social systems.* Lancaster, Pa.: Jacques Cattell Press, 1942.

179. Révész, G. The language of animals. *J. gen. Psychol.*, 1944, **30**, 117–147.

180. Riess, B. F. Freezing behavior in rats and its social causation. *J. soc. Psychol.*, 1946, **24**, 249–251.

181. Roberts, D. Imitation and sugges-

tion in animals. *Bull. Animal Behav. Lond.*, 1941, **1**, 11–19. (*Psychol. Abstr.*, 1942, **16**, No. 1396.)

182. Ross, S. Some observations on the lair dwelling behavior of dogs. *Behaviour*, 1950, **2**, 144–162.
183. Ross, S., & Ross, J. G. Social facilitation of feeding behavior in dogs: I. Group and solitary feeding. *J. genet. Psychol.*, 1949, **74**, 97–108.
184. Ross, S., & Ross, J. G. Social facilitation of feeding behavior in dogs: II. Feeding after satiation. *J. genet. Psychol.*, 1949, **74**, 293–304.
185. Ross, S., & Scott, J. P. Relationships between dominance and control of movement in goats. *J. comp. physiol. Psychol.*, 1949, **42**, 75–80.
186. Ross, S., Smith, W. I., & Denenberg, V. A preliminary study of individual and group hoarding in the white rat. *J. genet. Psychol.*, 1950, **77**, 123–127.
187. Royce, J. R. A factorial study of emotionality in the dog. *Amer. Psychologist*, 1950, **5**, 262. (Abstract)
188. Roubaud, E., & Provost, A. Influence des reactions d'ordre psychique chez certains animaux trypanosomes. *Bull. Soc. Path. Exot.*, 1940, **33**, 315–320. (*Biol. Abstr.*, 1941, **15**, No. 20190.)
189. Sabine, W. S. Dominance in winter flocks of juncos and tree sparrows. *Physiol. Zool.*, 1949, **22**, 64–85.
190. Schenkel, R. Expression studies in the wolf; captivity observations. *Behaviour*, 1947, **1**, 81–129.
191. Schneirla, T. C. Social organization in insects, as related to individual function. *Psychol. Rev.*, 1941, **48**, 465–486.
192. Schneirla, T. C. Contemporary American animal psychology in perspective. In P. L. Harriman (Ed.), *Twentieth century psychology.* New York: Philosophical Library, 1946. Pp. 306–316.
193. Schneirla, T. C. The relationship between observation and experimentation in the field study of behavior. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1022–1044.
194. Schneirla, T. C. A consideration of some problems in the ontogeny of family life and social adjustments in various infra-human animals. In M.E.J. Senn (Ed.), *Problems of infancy and childhood.* New York: Josiah Macy, Jr. Foundation, 1951. Pp. 81–124.
195. Scott, J. P. Hereditary differences in social behavior (fighting of males) between two inbred strains of mice. *Anat. Rec.*, 1940, **78**, 103.
196. Scott, J. P. Genetic differences in the social behavior of inbred strains of mice. *H. Hered.*, 1942, **33**, 11–15.
197. Scott, J. P. Social organization and leadership among sheep. *Anat. Rec.*, 1942, **84**, 480–481.
198. Scott, J. P. Social organization in a small flock of domestic sheep. *Proc. Indiana Acad. Sci.*, 1942, **51**, 261.
199. Scott, J. P. Age as a factor affecting allelomimetic social dominance in a small flock of sheep. *Anat. Rec.*, 1943, **87**, 471–472.
200. Scott, J. P. Difference in social organization of mice caused by difference in social behavior (fighting of males). *Genetics*, 1943, **28**, 88–89.
201. Scott, J. P. Social organization and leadership among sheep. *Proc. Indiana Acad. Sci.*, 1943, **52**, 187.
202. Scott, J. P. Social behavior, range and territoriality in domestic mice. *Proc. Indiana Acad. Sci.*, 1944, **53**, 188–195.
203. Scott, J. P. An experimental test of the theory that social behavior determines social organization. *Science*, 1944, **49**, 42–43.
204. Scott, J. P. Social behavior, organization and leadership in a small flock of domestic sheep. *Comp. Psychol. Monogr.*, 1945. **18**, 1–29.

205. Scott, J. P. Group formation determined by social behavior; A comparative study of two mammalian societies. *Sociometry*, 1945, **8**, 42–52.

206. Scott, J. P. Dominance reactions in a small flock of goats. *Anat. Rec.*, 1946, **94**, 380–381.

207. Scott, J. P. Incomplete adjustment caused by frustration of untrained fighting mice. *J. comp. Psychol.*, 1946, **39**, 379–390.

208. Scott, J. P. "Emotional" behavior of fighting mice caused by conflict between weak stimulatory and weak inhibitory training. *J. comp. physiol. Psychol.*, 1947, **40**, 275–282.

209. Scott, J. P. Dominance and the frustration aggression hypothesis. *Physiol. Zool.*, 1948, **21**, 31–39.

210. Scott, J. P. Studies on the early development of social behavior in puppies. *Amer. Psychologist*, 1948, **3**, 239–240. (Abstract)

211. Scott, J. P. Genetic differences in social behavior of dogs. *Amer. Psychologist*, 1950, **5**, 261. (Abstract)

212. Scott, J. P. (Ed.) Methodology and techniques for the study of animal societies. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1001–1122.

213. Scott, J. P. The social behavior of dogs and wolves: An illustration of sociobiological systematics. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1009–1021.

214. Scott, J. P., & Marston, M. V. Critical periods affecting the development of normal and maladjustive social behavior of puppies. *J. genet. Psychol.*, 1950, **77**, 25–60.

215. Scott, J. P., & Marston, M. V. The development of dominance in litters of puppies. *Anat. Rec.*, 1948, **101**, 696.

216. Scott, J. W. A study of the phylogenetic or comparative behavior of three species of grouse. *Ann. N. Y. Acad. Sci.*, 1950, **51**, 1062–1073.

217. Seward, G. H. The role of hunger in the filial behavior of the young guinea pig. *Psychol. Bull.*, 1939, **36**, 563. (Abstract)

218. Seward, J. P. Aggressive behavior in the rat. I. General characteristics; age and sex differences. *J. comp. Psychol.*, 1945, **38**, 175–197.

219. Seward, J. P. Aggressive behavior in the rat. II. An attempt to establish a dominance hierarchy. *J. comp. Psychol.*, 1945, **38**, 213–224.

220. Seward, J. P. Aggressive behavior in the rat. III. The role of frustration. *J. comp. Psychol.*, 1945, **38**, 225–238.

221. Seward, J. P. Aggressive behavior in the rat. IV. Submission as determined by conditioning, extinction and disuse. *J. comp Psychol.*, 1946, **39**, 51–76.

222. Shadle, A. R. The play of American porcupines. *J. comp. Psychol.*, 1944, **37**, 145–150.

223. Shlaifer, A. An analysis of the effect of numbers upon the oxygen consumption of *Carassius auratus*. *Physiol. Zool.*, 1939, **12**, 381–392.

224. Shlaifer, A. The locomotor activity of the goldfish, *Carassius auratus L.*, under various conditions of homotypic and heterotypic grouping. *Ecology*, 1940, **21**, 488–500.

225. Shlaifer, A., & Breder, C. M., Jr. Social and respiratory behavior of small tarpon. *Zoologica* (N. Y.), 1940, **25**, 493–512.

226. Smith, E. M. A study of the development of social behavior in pups. Bar Harbor: R. B. Jackson Memorial Laboratory; unpublished. (Cited by Scott, **210**).

227. Spence, K. W. Réaction des mères chimpanzes à l'égard des enfants chimpanzes après séparation. *J. Psychol. norm. path.*, 1937, **34**, 475–494. (*Psychol. Abstr.*, 1938, **12**, No. 4033.)

228. Stewart, J. C., & Scott, J. P. Lack of correlation between leadership

and dominance relationships in a herd of goats. *J. comp. physiol. Psychol.*, 1947, **40**, 255–264.

229. TINBERGEN, N. On the analysis of social organization among vertebrates, with special reference to birds. *Amer. Midland Nat.*, 1939, **21**, 210–234.

230. TINBERGEN, N. Die Übersprungbewegung. *Z. Tierpsychol.*, 1950, **4**, 1–40. (Cited by Tinbergen, **228**.)

231. TINBERGEN, N. An objectivistic study of the innate behavior of animals. *Bibliotheca biotheoretica*, Series D, 1942, **1**, 39–98.

232. TINBERGEN, N. Social releasers and the experimental method required for their study. *Wilson Bull.*, 1948, **60**, 6–51.

233. TINBERGEN, N., & KUENEN, D. J. Über die Auslösenden und die richtunggebenden Reizsituationen der Sperrbewegung von jungen Drosseln. *Z. Tierpsychol.*, 1939, **3**, 37–60.

234. TINKLEPAUGH, O. L. Social behavior of animals. In F. A. Moss (Ed.), *Comparative psychology.* (Rev. Ed.) New York: Prentice-Hall, 1946. Pp. 366–293.

235. TSAI, L. S. Rivalry and cooperation in white rats. *Amer. Psychologist.* 1950, **5**, 262. (Abstract)

236. UHRICH, J. The effect of experience on fighting behavior of albino mice. *Ecology*, 1940, **21**, 100–101.

237. VOGEL, H. H., JR. Social relations among Springer spaniel dogs. *Anat. Rec.*, 1947, **99**, 625.

238. VORONIN, L. G. Kvoprosu ob imitatsionnykh sposobnastiakh u nizhikh obez ian. *Fizial. Zh. S.S.S.R.*, 1947, **33**, 373–380. (*Psychol. Abstr.*, 1950, **24**, No. 1052.)

239. WALKER, E. P. Animal behavior. *Ann. Rept. Smithsonian Inst.*, 1941, 271–312. (*Biol. Abstr.*, 1942, **16**, No. 3663.)

240. WARDEN, C. J., FJELD, H. A., & KOCH, A. M. Imitative behavior in cebus and rhesus monkeys. *J. genet. Psychol.*, 1940, **56**, 311–322.

241. WARDEN, C. J., & GALT, W. A. A study of cooperation, dominance, grooming, and other social factors in monkeys. *J. genet. Psychol.*, 1943, **63**, 213–233.

242. WARNE, M. C. A time analysis of certain aspects of the behavior of small groups of caged mice. *J. comp. physiol. Psychol.*, 1947, **40**, 371–387.

243. WHITE, L. On the use of tools by primates. *J. comp. Psychol.*, 1942, **34**, 369–374.

244. WINSLOW, C. N. A study of experimentally induced competitive behavior in the white rat. *Comp. Psychol. Monogr.*, 1940, **15**, No. 6, 1–35.

245. WINSLOW, C. N. Patterns of competitive, aggressive, and altruistic behavior in the cat. *Psychol. Bull.* 1941, **33**, 564. (Abstract)

246. WOLFLE, D. L., & WOLFLE, HELEN M. The development of cooperative behavior in monkeys and young children. *J. genet. Psychol.*, 1939, **55**, 137–175.

247. WOODBURY, A. M. Changing the "hook order" in cows. *Ecology*, 1941, **22**, 410–411.

248. YERKES, R. M. Social dominance and sexual status in the chimpanzee. *Quart. Rev. Biol.*, 1939, **14**, 115–136.

249. YERKES, R. M. Social behavior of chimpanzees: Dominance between mates in relation to sexual status. *J. comp. Psychol.*, 1940, **30**, 147–186.

250. YERKES, R. M. *Chimpanzees: A laboratory colony.* New Haven: Yale Univer. Press. 1943.

Received February 13, 1952.

Hull, Clark L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951. Pp. viii+145. $2.75.

The eight years between the publication of *Principles of Behavior* and *Essentials of Behavior* have been busy ones in the field of learning and conditioning. New experimental data have been accumulating at an impressive rate, at the hands of both Hull and his followers, who were testing the system, and other researchers who were working without reference to the way in which their studies might bear on Hull's system. In the face of this new knowledge, many of us have wondered what changes had occurred in Hull's thinking over these years. The present book is the answer, or at least part of the answer.

This is a little book, scrupulously stripped down to a crisp and pithy presentation of its core ideas. It offers a restatement of the earlier system of the *Principles*, with the embedding contextual material reduced to a near minimum. The author says: "During the last eight years the fundamental hypotheses of my system have matured considerably. This development will be reflected in the system itself, which I hope to publish within a year or two" (p.v). Hull was responsive enough to the new data to feel that public emendation of his earlier system was called for early, though he apparently had a more thorough recasting already in mind.

Besides containing the latest statement of Hull's theory of behavior, it was the author's hope that the present book would be both "intelligible" and "a useful introduction to the current aspects of behavior theory." Newcomers to Hull, however, will probably find that the conciseness of the book is gained at the cost of intelligibility. All theories require a certain amount of plain language context to be comprehensible. This little book does not have enough, and a reader must be prepared to draw upon the context of Hull's earlier writings, particularly the *Principles*. This will be no hardship for the cognoscenti, who simply want Hull-to-date, and who may welcome the brevity of the book; but others, not yet familiar with Hull, will need to learn the earlier work before they can assimilate this latest one. For the same reason, and apart from any questions of the systematic adequacy of the theory itself, the book will not serve for newcomers the introductory purpose which its author had in mind.

The similarities between the *Principles* and the *Essentials* are deep-going, and compose a fundamental continuity between the two books. Here is the same philosophy of science, the same theoretical orientation to psychology, the same approach to theory construction. Hull's version of postulate method, still in force, seems to have undergone no refinement since the *Principles*, just as his faith in it has undergone no weakening. Curve-fitting is still used, in place of rational equations, not only for summarizing a plot of empirical data, but for isolating many of the postulates entering into the theory. The system, in many respects, is still much as it was, in substance as well as form. It takes its departure from concepts like need and survival; the same variables are featured; the same few types of arbitrary functional relations do most of the work; some of the old postulates reappear almost intact. Hull is still preoccupied with quantification because his belief is unshaken that only in that way can "scientific" progress be made by behavior theory. Comments on these aspects of Hull's theory may be found in previous reviews of the *Principles* and in critical papers on Hull's general system and methodology.

Of more current interest are the changes that mark this new version of Hull's system. A few of the broader differences may be cited to show the scope of those changes. Hull says, " . . . numerous reasons for changes and modifications in the postulates as presented [in the *Principles*] have been revealed." Accordingly, he continues, " . . . the mathematical aspects of many of the postulates have been formulated, or reformulated, and the verbal formulation of nearly all has been modified to a certain extent. One postulate . . . has been dropped in part as empirically erroneous; some postulates have been divided and others have been combined; several new postulates have been added; and a number of the original postulates have been deduced from others of the present set and appear as corollaries in these pages." All this reshuffling has led to considerable trimming down in the total number of postulates, subpostulates, major corollaries, and minor corollaries to only a little more than half their number in the *Principles*. Part of this reduction, it may be noted, is achieved by omitting certain topics like stimulus patterning and *S-R* asynchronism. There has been a shift in emphasis from ${}_SH_R$ to ${}_SE_R$, the latter being described now as "the central quantitative concept of the theoretical system," and an attempt is made several times to rationalize this shift. Much attention is cast upon the problem of reaching a quantification of "true" ${}_SE_R$, with Thurstone's Case III taken as the key procedure; and the concept of probability of response is applied to the interpretation of ${}_SE_R$. With regard to basic viewpoint, too, the *Essentials* offers some interesting compromises and corrections. Thus, to criticisms of his physiologizing in the *Principles*, Hull replies that a term like afferent neural interaction is "at bottom purely molar," and he points to the "methodology put forward for quantifying afferent stimulus interaction" as proof of its molar character. Or, again, he addresses himself anew to the question of whether there are one or two types of conditioning, reasserting his belief that there is only one on his customary contention that reinforcement or drive-reduction is involved in all cases. (The term "law of effect" is not mentioned once.) At the same time, he compromises by harking back, and giving weight, to an issue which figured in the Skinner and Konorski-Miller discussion, namely, the *de novo* character of CR's obtained with one conditioning procedure and not with the other.

Comparisons such as we have been making between the *Principles* and *Essentials* are, however, a matter of historical interest mainly. What is more important is that we have in the latter a system of behavior brought up to date, presumably complete in itself, and calling for a critical inspection without reference to its earlier forms. To make this inspection fully in our space is impossible, but it will perhaps be worth while to report briefly on one part of the theory as an example of his treatment. For this, the handling of ${}_sE_R$, and the variables constituting ${}_sE_R$, may be chosen.

As in the *Principles*, the necessary and sufficient conditions for the acquisition of new *S-R* connections or for the strengthening of old ones is the occurrence of reinforcement. Reinforcement is said to involve drive stimulus reduction rather than need reduction. Nevertheless, it is clear from Hull's remarks (e.g., p. 15) that the development of a state of need is still considered essential, since it underlies drive and the consequent production of drive stimuli. Hull appears to be as far as ever from an operational definition of reinforcement. Postulate III, which embodies the "law of reinforcement," fails to clarify the matter. It includes a mixture of terms dealing with observable (e.g., Response) and unobservable (e.g., S_D, s) events and specifies only that drive stimulus reduction must be "rapid." Obviously, this qualification leaves much room for ad hoc explanations of the failure of an apparently "reinforcing state of affairs" to strengthen behavior, since the hypothesized drive stimulation is internal, and no rules are supplied for measuring it, let alone the rapidity of its diminution.

${}_sH_R$, which is generated by the reinforcement process, and drive (D) are the key variables which determine reaction potential (${}_sE_R$). As stated earlier, ${}_sE_R$ is the "central theoretical construct" of the system, but it is at the same time asserted that the procedure involved in the quantification of ${}_sE_R$ constitutes its operational definition. Since this quantification is specifically based upon latency measurements, this makes ${}_sE_R$ essentially a derived response measure rather than a theoretical construct. Thus, in view of the above assertion, equations relating ${}_sE_R$ to theoretical variables like ${}_sH_R$, J, D, etc. might be converted to the form $R=f$ (intervening variables), while the equations relating response measures like amplitude to ${}_sE_R$ might take the form $R=f(R)$.

Hull's procedure in handling ${}_sE_R$ is the same as that utilized in the *Principles*. Explicit functions relating variables like ${}_sH_R$, D, J, etc. to environmental factors are written first. Then ${}_sE_R$ is related to the intervening variables by general equations stated in functional notation form [e.g., ${}_zE_R=f(D\times{}_sH_R)$], though Hull sometimes omits the "*f*" symbols (cf. equation 19, p. 59). Finally, ${}_SE_R$ is tied up, again by explicit functions, with such response measures as latency, amplitude, number of unreinforced responses in extinction, etc. It may be noted incidentally that the latter equations, written in the form ${}_SE_R=f(R)$, would be more appropriate in the form $R=f({}_SE_R)$, since the amplitude, latency, etc. of response presumably depend upon ${}_SE_R$ as determined by the values of ${}_SH_R$, D, etc., and, in the final analysis, by the values of N, h, w, etc., the antecedent empirical variables. Special status is, however, given to latency measurements through their use in the quantification or operational definition of ${}_SE_R$.

Of the intervening variables upon which ${}_SE_R$ depends, ${}_SH_R$, V_1, K and J are functions of conditions at the time of acquisition, whereas D and V_2 are considered to exert their influence during performance and extinction. That D plays its role during performance or extinction is an inference based upon Hull's derivation of the basic equation $D=f(h)$ from extinction data supplied by Yamaguchi. However, the matter is somewhat confused by the wording of Postulate IX (p. 59) to the effect that "the reaction potential . . . is determined (1) by the drive (D) operating *during the learning process* . . . ." (Our italics.)

As noted earlier, the most important acquisition variable determining ${}_SE_R$ is ${}_SH_R$. According to Hull, ${}_SH_R$ depends primarily upon N, the number of reinforcements. The so-called "law of habit formation" deals with the relation ${}_SH_R=f(N)$. The latest version of this law appears to be more restricted than the one presented in the *Principles of Behavior*. It now deals with reinforcements occurring at "evenly distributed intervals" (cf. Postulate IV, p. 32). At least three points should be noted concerning Postulate IV: (*a*) the law omits the case of unevenly spaced reinforcements; (*b*) partial or intermittent reinforcement continues to be ignored; (*c*) neither in this postulate nor elsewhere in the volume is there any systematic mention of spacing as a parameter of acquisition.

J and K, like ${}_SH_R$, play a new role in the current version of the system. Instead of influencing the upper limit of ${}_SH_R$ as a function of N, they have become direct determiners of ${}_SE_R$. As was true for ${}_SH_R$, the postulate relating K to its empirical variable is much more restricted than the equivalent statement in *Principles*. Postulate VII (p. 51) deals with K merely as a function of w, the weight (gm) of food. It fails to consider non-food reinforcement, or the quality of reinforcement, both covered by the earlier postulate. The derivation of the equation relating K to w is based primarily upon data supplied by Crespi. While admitting that only three of the Crespi curves were sufficiently regular to permit empirical curve fitting, Hull proceeds to use asymptote values based upon five. He further omits the asymptote value for the zero incentive curve on the ground that it is "presumably distorted upward by frustration" (p. 49). It is then assumed, without further explanation, that the three highest data points graphed in Fig. 12 correspond to the coefficients of the empirical equations fitted to the three suitable Crespi curves, "except that they all have considerably smaller values."

In deriving the delay of reinforcement equation (Postulate VIII, p. 56), there is again some tendency to blink the data. When equation 15, basic to Postulate VIII, is fitted to the data provided by Perin's study, the resulting curve indicates an asymptote at an above-zero value of $_SE_R$ (see Fig. 14, p. 55). Yet in the Perin study, animals failed to condition at a 30-sec. delay, denoting zero value of $_SE_R$ at that delay. This indicates either the existence of a discontinuity requiring a second function starting at some value above 10 sec. (the last plotted point), or at least the need for a more complex function, but the discrepancy is dismissed as "probably due to some artifact."

Among the intervening variables which together constitute reaction potential are two which do not appear in any form in the earlier volume. These are V_1 and V_2, stimulus intensity dynamism variables for acquisition, and for performance ("response evocation") or extinction, respectively. That is, the intensity of the stimulus at the time of conditioning, and during later response evocation, is considered to affect response strength apart from all other variables. V_2, the performance variable, affects $_SE_R$ directly, being given equal status with J, K, etc. V_1, on the other hand, is considered to affect $_SH_R$, and a new symbol, $_S\underset{\cdot}{H}_R$, is created to represent the resultant habit strength. But, according to equation 10′ (p. 45), $_S\underset{\cdot}{H}_R = {}_SH_R \times V_1$, and according to equation 19 (p. 59), $_SE_R = D \times V_2 \times \cdots \times {}_S\underset{\cdot}{H}_R$. Thus, by substitution, we get $_SE_R = D \times V_2 \times \cdots \times {}_SH_R \times V_1$, and the reason for the different status of V_1 and V_2, or for the creation of the symbol $_S\underset{\cdot}{H}_R$ becomes difficult to see.

In the attempt to demonstrate the need for stimulus intensity variables, Hull leans heavily upon an unpublished study by Hays (pp. 41–43). Apart from the admitted irregularity of the data for the group jumping to a white target, no reason is offered for the fact that the group jumping to a black target should have demonstrated virtually no conditioning in spite of 143 reinforcements spread over 13 days, and the reader is left to conjecture what the cause of the demonstrated difference might be. Whatever the cause, the study appears to provide a rather tenuous basis for the introduction of new intervening variables. Several further thoughts must occur to the reader regarding the theoretical constructions surrounding V_1 and V_2.

1. From the equations fitted to the Hays data, the expression $_SE_R = f(V_2 \times {}_SH_R)$ is derived. But V_2 applies to "response evocation," while V_1 acts during "original learning." There seems to be no clear reason for treating the Hays study as being of the former case. Furthermore, the problem of deciding when "original learning" ends and "response evocation" begins would seem to be a difficult one at any time.

2. V_1 is introduced in equation 10′ without any attempt to demonstrate the need for such a variable at either the experimental or the theoretical level. (As noted above, Hays' study is used in connection with V_2.)

3. The assumption concerning Cattell's unit (p. 44), used to derive equation 10, seems completely arbitrary. This, together with other similar instances in the book, makes the reader pause to consider the general question of the legitimacy of making assumptions, however arbitrary, because they lead to equations which fit well data for whose fit the assumptions were created.

4. The absence of separate equations relating V_1 and V_2 to the empirical variable, *S*, presumably means that the equation in Postulate VI holds for both cases, but this is never stated explicitly and the reader is left to make his own inference.

The final intervening variable, Drive (*D*), which contributes to $_SE_R$, is sometimes considered by Hull to be the most important of all (*vide*, p. 58). As noted earlier, *D* presumably refers to drive at the time of performance or extinction. However, some above-zero value of drive is required also during conditioning, on Hull's assumption that reinforcement is a drive (or drive stimulus) reduction process. Some lack of clarity is evident in the discussion of drive. At times drive is described as a form of stimulation (p. 15); at other times as a need (pp. 3–4, 15); at still other times as a state (p. 15). Possibly the author intends these terms to be taken as synonymous, but there are occasions when they seem to have separate referents (e.g., p. 39, legend of Fig. 4, p. 18), and the matter is never quite settled. In fact, on p. 76, drive reduction is referred to as only "*one* of the necessary conditions of reinforcement." (Our italics.)

In spite of such verbalizations, one role of drive in the system is unmistakably that of an intervening variable (*vide*, p. 33). As was the case with $_SH_R$ and *K*, the drive postulate is quite restricted in scope. With the exception of parts *C* and *D*, it deals solely with the hunger drive. Parts *A* and *B* of the postulate, as well as equation 7, are taken directly from data supplied by Yamaguchi. The latter investigator derived an equation relating $_SE_R$ to *h*, hours of food deprivation. In the book the same expression appears as equation 7, but relates *D*, not $_SE_R$, to *h*. Such a substitution would be justified only if *D* and $_SE_R$ were the same, or if *D* were a close approximation of $_SE_R$. The tendency to speak of drive as though it were some form of reaction potential is further illustrated by the wording of Postulate V, part B (pp. 38–99). The $_SE_R$ values computed by Yamaguchi from his extinction data are graphed (Fig. 8, p. 36), and equation 7 is fitted to them. While the curve as drawn passes through a maximum, the trend of the values seems to provide as good a basis for fitting a curve rising toward an asymptote. The fact that the plotted curve fits the data closely means little, since its equation has as many constants as there are computed values. (On the basis of equation 7, Hull follows Yamaguchi's lead in adding to the effects of food deprivation an inanition component ϵ

trary reasons, and then only on one side of the equation. On p. 81, n is set equal to two different expressions in equations 33 and 34. Then the right side of equation 34 is substituted for one constant in equation 33. Typographical errors are rare in the book: one case is found on page 94 where the equation 25 referred to deals with K and not V.

Despite the brevity of this little book, it unmistakably carries Hull's personal seal and flavor. On occasion, he steps outside the limits of presenting his theory; here, to expound on scientific method; there, to exhort his reader to the "fearless performance of critical experiments"; and elsewhere, to give us his own chronology of growth as a behavior theorist, "through twenty-five volumes of hand-written notebooks," beginning in 1915 (or three years before he received his doctorate). Stylistically, too, the prose identifies itself as Hull's familiar hand.

Hull's theoretical work has always been marked by an effort to achieve explicitness and testability, because he believed those virtues to be essential to progress in theory construction. They also make criticism easier, as Hull was well aware, but he properly regarded that as a requisite of progress. The *Essentials* was his attempt to improve on the *Principles*, perhaps only in a stopgap way, since in it he declared his intention soon to publish a major work. This last project has now been cut short by his untimely death. Even his critics will miss that next book.

W. N. SCHOENFELD.
P. J. BERSH.

Columbia University.

PARSONS, TALCOTT, & SHILS, EDWARD A. (Eds.) *Toward a general theory of action.* Cambridge: Harvard Univer. Press, 1951. Pp. xi+506. $7.50.

This volume is the result of a project, subsidized by the Carnegie Corporation and the Harvard Department of Social Relations, designed to assess "the theoretical foundations underlying the synthesis which had been worked out on the organizational level through the foundation of the Department of Social Relations." As such, it becomes an important test case of the fruitfulness of the "interdisciplinary" attack on problems of social science theory. Since the furious waving of the interdisciplinary banner in recent years has mobilized many recruits and strong institutional support, the volume deserves careful study, and certainly far more detailed analysis than can be given it in this brief review.

The project staff consisted of two sociologists (Talcott Parsons and Edward A. Shils), a psychologist (Edward C. Tolman), and a social anthropologist (Richard C. Sheldon). During the five-month period of the project, the staff operated both as an independent unit and in conjunction with two secondary echelons—one, a discussion group including G. W. Allport, C. Kluckhohn, H. A.

Murray, R. R. Sears, and S. A. Stouffer, and the other, a larger discussion group including "all interested members of the Department of Social Relations." Funneling from this pool of interdisciplinary effort into the present book is a brief general statement on "Some Fundamental Categories of the Theory of Action," drafted by Parsons and signed by all contributors; a lengthy elaboration of the theory in the form of an essay on "Values, Motives, and Systems of Action" by Parsons and Shils (with the assistance of James Olds); the specification of "A Psychological Model" by Tolman; and a group of briefer contributions on "applications" of the theory of action by the five secondary participants named above. Sheldon contributes a short methodological section, "Some Observations on Theory in the Social Sciences." In violation of the spirit of interdisciplinary fair-play, it may be revealed that the formulations of Parsons and Shils dominate the book both in the allotment of space and conceptually. Accordingly, they must dominate this review.

The "theory of action" is mainly a set of formulations which talk about the variables of a *projected* theory. There is also, as is becoming inevitable in social science, much talk about the talk about the variables of this projected theory. Beginning with this second level of talk (the "meta-meta-language?"), we itemize a few of the authors' conceptions about the objectives and properties of the theory of action:

1. The aim is to "develop a unified conceptual scheme for theory and research in the social sciences."
2. In the execution of this task, the authors propose to do justice to the "whole course of development of work in the social sciences" (p. 239). In particular, they claim responsiveness to the emphases of Freud, Allport, and the study of animal behavior (as mediated by Tolman) in psychology; in sociology, to Durkheim, Weber and, to a lesser extent, Park and Thomas; in anthropology, to Boas, Kroeber, Sapir, and others.
3. The *immediate* intention is to identify a set of variables adequate to the "descriptive analysis" of the state of "systems of action" at a given moment. However, these variables are to be chosen in such a way as to permit (ultimately) "dynamic analysis" of the "changes of the system through time," in terms of changes in the values of the variables.
4. The authors distinguish four "levels" of scientific systematization: "(1) *ad hoc* classificatory systems, (2) categorical systems, (3) theoretical systems and (4) empirical-theoretical systems" (p. 50). The "categorical type . . . involves a system of classes which is formed to fit the subject matter, so that there are intrinsic relations among the classes, and these are in accord with the relations among the items of the subject matter" (p. 50). A "theoretical system is a categorical system whose laws relating elements [i.e., the categorical variables or "classes"] have been formulated." Theoretical systems generate deductive consequences which "make . . . predictions about . . . how empirical systems . . . behave if an ideal set of scientific or 'standard' conditions were to exist" (p. 51). "Empirical-theoretical systems . . . the long term goal of scientific endeavor" . . . are capable of "predicting changes in empirical systems outside special experimental conditions" (p. 51). Parsons and Shils locate the present status of the theory of action in the *categorical* domain.

5. Nevertheless, despite the fact that their categorical system is admittedly "free of laws which relate its elements," the authors claim, at many points, to be using *deductive* techniques, and, in general, feel that they "carry deductive procedures further than is common in the social sciences" (p. 49).

Turning now to the talk about the variables of the proposed theory of action, we face a formidable problem. So fully do the "categories" of the theory reflect the complexities of its subject matter, that a meaningful précis is impossible. Indeed the some 250 closely printed pages devoted to the theory is itself a précis, consisting of a relentless barrage of definitions and classificatory elaborations of these definitions, but devoid either of empirical application or illustrative explication of the defined variables. Perhaps the following paragraphs will impart the flavor, if not the sense, of the formulations.

The theory of action makes statements about the action of individual *actors* or collectivities of actors. The descriptive unit for the analysis of action is the *orientation of action*. An orientation of action "involves a set of *objects of orientation*," characteristic of the *situation*. The two principal classes of objects provided by the situation are: "(1) nonsocial, that is, physical objects or accumulated cultural resources, or (2) social objects, that is, individual actors and collectivities" (p. 5). "A specific combination of selections relative to . . . objects," made from the available possibilities of selection in a specific situation, defines an *orientation of action* in a given case. The orientation of action may be broken down into two analytically independent sets of categories: *motivational orientation*, and *value-orientation*. "Motivational orientation refers to those aspects of the actor's orientation . . . which are related to actual or potential gratification of the actor's need-dispositions" (p. 58) and is characterized by three *modes: cognitive, cathectic*, and *evaluative*. "The *cognitive* mode involves the various processes by which an actor *sees* an object in relation to his system of need-dispositions." "The *cathectic* mode involves the various processes by which an actor invests an object with affective significance." "The *evaluative* mode involves the various processes by which an actor allocates his energy among the various actions with respect to various cathected objects in an attempt to optimize gratification." "*Value-orientation* refers to those aspects of the actor's orientation which commit him to the observance of certain norms, standards, criteria of selection, whenever he is in a contingent situation which allows . . . him to make a choice" (p. 59). Such norms or standards exist as *need-dispositions* within the actors who observe these rules. When abstracted from the concrete value-orientations of actors, these rules form a set of cultural *value standards*. Three *modes* of value-orientation—*cognitive* (standards of validity), *appreciative* (standards of "immediate gratificatory significance," e.g., aesthetic), and *moral* (standards which assess the consequences of actions for the "integration" of personality, and the "integration of the social systems" in which the actor participates)—"parallel the modes of motivational orientation" (p. 60).

An "organized plurality of . . . orientations of action constitutes a system of action" (p. 5). The "organized system of the orientation and motivation of action of one individual actor" is the *personality system*. A *social system* is a system of the "interactions of a plurality of individual actors" (p. 7). *Cultural systems* are not, strictly speaking, systems of *action*. Culture exists "as a body of artifacts and as systems of symbols," and is embodied in the "orientation systems of concrete actors" (p. 7). The variables for the description of personality systems derive from various "fundamentals of behavior psychology," from psychoanalysis and other sources, and are brought to converge on the concept of *need-dispositions* as the basic organizational unit. *Role* is the unit of the social system, while cultural systems are *systems of ideas*, *systems of expressive symbols*, and systems of *standards of value-orientation*. Despite the existence of complex interrelations, and a homology of certain of the concepts at the various levels, social systems are not reducible to personality systems, nor are cultural systems reducible to any combination of the variables of either. A major concern of the theory of action is to locate certain "points of articulation" among the three types of "systems," a pivotal such area being the concept of *role*, where "value pattern, social structure, and personality come together" (p. 243).

Wherever one moves in the theory of action, one stumbles into a set of five classificatory dichotomies called "the pattern variables," to wit: affectivity—affective neutrality; self-orientation—collectivity-orientation; universalism—particularism; ascription—achievement; specificity—diffuseness. The "pattern variables" had originally been created by Parsons for the classification of certain structural components of *social* systems. In the over-all theory of action, it happily turns out that the pattern variables are capable of momentous feats of integration at every level. They serve as the basis for characterizing the actor's concrete *value-orientations*. They lead, via 32 cross classifications, to the cataloging of *need-dispositions*. Similarly, they lead, on the level of social system, to a classification of *role-expectations*. They enter also "on the cultural level as aspects of value standards" (p. 78).

The mode of presentation of the theory of action imposes severe difficulties on any attempt to characterize its methodological status. Its authors write as if fighting a continuous but losing battle to express the ineffable. The exposition consists of a series of definitions in abstract, discursive verbiage. Each concept is defined not once but in many different contexts throughout the discussion, so that a given term acquires a complex and slippery array of meanings by a process of cumulative accretion. It is hard to find a place for many of the definitional statements within the logic of definition. The authors' belief that they are using *deductive* procedures seems to spring, in part, from the intention to construct explicit definitions from the so-called "elements" (primitive terms) of the "frame of reference of the theory of action." However, these "elements" or primitives—which strictly speaking require operational and perhaps implicit definition, but may not be explicitly defined—are given explicit definitions, often in terms of the later, presumably derivative, concepts.

The authors show laudable modesty in classifying the "theory of action" as a "categorical system." Nevertheless, it remains impossible to evaluate the potential *theoretical* fruitfulness of the variables ("categories") put forward. All that the "system" contains is an extended set of explicit definitions. Operational (empirical) definitions are not given for a single one of the many concepts, nor are there clues in the discussion as to how and from what sources such definitions might be constructed. Sheldon, in his contribution on the methodology of social science theory, notes this embarrassing circumstance, but derives much solace from the belief that theoretical concepts (and principles) are "free creations of the human intellect." Whatever the virtues of Poincaréan "conventionalism," it always remains desirable to ask, "how free?" And the wisdom of constructing an intricate array of theoretical concepts independent of step-by-step efforts, however tentative, to tie these concepts to empirical states of affairs may be seriously doubted. The theory of action does not even permit the determination of *where* operational definitions are relevant, because few grounds are available for deciding which of the "categories" correspond to systematic independent variables, which to intervening variables, and which to dependent variables.

Tolman's section on "A Psychological Model" marks one of the frequent new phases in his thinking. Tolman now feels that all intervening variables used in psychological theory are defined *both* functionally, and by the attribution of "postulated, ostensive [i.e., "existential"] properties." The referents may be "neurophysiological," "phenomenological," or the components of "*a sui generis model*." Tolman's new theory develops such a model "by various analogies from simple physics and mechanics, from Lewin's 'topological and vector' psychology, and from common experience" (p. 283). The attempt to do justice to certain of the conceptual demands of the theory of action unleashes the Lewin in Tolman to a greater extent than ever before. Indeed, the result is very much like a reshuffling of Lewin into a forced independent-intervening-dependent variable schema, with the addition of a few special twists to accommodate some of the requirements of the theory of action—all this decorated with many Tolmanesque flourishes.

Tolman's model relates environmental, "drive arousal," and individual difference, independent variables to a complex set of intervening constructions. The major intervening variable units are: (*a*) a *need system* (Lewin's treatment of "need structure" ripped out of the life-space and massaged a bit by Tolman), (*b*) a *belief-value matrix* (the repository for Parson's "need-dispositions" and Tolman's former "means-end-readinesses"), (*c*) an *immediate behavior space* (Lewin's momentary life-space as modified to contain Tolman's "expectations") and (*d*) a *restructured behavior space* (the "immediate" space as "restructured" by "locomotions," "learning," or the "psychodynamic mechanisms"). The postulated interrelations among these "variables" (and the subvariables constitutive of each) are of the customary programmatic or "need to know" sort. At many points, the reliance on Lewin is more a matter of form than of content, for many of the apparently Lewinian concepts (e.g., "distance," "direction," "locomotion") are homonyms used in quite a different sense. The model is also endowed with "'electromagnetic' charges" as needed, and is often the victim of such incredibly ad hoc calumnies as: "The *strength* of a field force is . . . directly proportional to the product of the need-push and the determining valence . . . and inversely proportional to the square of the behavior-space distance between the region of the behaving self . . . and the region of the corresponding valence" (p. 340).

The model is indeed *sui generis*. Tolman, however, berates the model with his usual candor, often implying that it may well prove necessary to do away with the thing. Should this be the model's unhappy fate, the field will nevertheless recall with nostalgia a fetching feature of its conceptual physiognomy called "the nose of the behaving self."

Of the five "applications" of the theory of action which comprise the final portion of the book, only two are related to the theory in any determinate sense. One of these, Kluckhohn's "Values and Value-Orientations in the Theory of Action," is not an *application*, but a further definitional analysis of this topic. The other, "An Empirical Study of Technical Problems in Analysis of Role Obligation" by Stouffer, concludes with this warning: "one of the most important values of this paper should be its service as a brake on the enthusiasm of those who may anticipate quick and easy progress in moving from highly abstract concepts of social science to empirical operations" (p. 494). Murray's essay on the classification of needs, and Sears' paper on the desirability of dealing with individual and social behavior within a single theory, are both relevant to the theory of action in some broad sense, but do not proceed within the "categories" of the theory. Perhaps the most astounding property of the theory of action is its capacity to "postdict," as an *application*, Allport's 1950 SPSSI address on "Prejudice."

At a time when the route towards, *even the possibility of*, comprehensive theory in each of the individual social sciences is in doubt, it is more than sanguine to begin the construction of a unified theory for *all* social science. The "theory of action," in its present form, gives substantial support to this statement. Yet, it is possible to argue that certain of the limitations of attempts at theory in the individual social sciences may derive from failure to explore concrete interdisciplinary relations among emerging concepts and empirical findings. Psychological theorists who try to deal with complex aspects of human behavior have been too long content to talk in such diffuse terms as "the role of *social* or *cultural factors*" within the individual behavior phenomena that they analyze, or "the effects of the *social environment*." It is possible that each one of the social sciences could benefit from modest attempts towards the coordination of developing empirical knowledge, and mutual accommodations among concepts. Embedded in the ponderous formulations of the "theory of action" are evidences of the possible fruitfulness of interdisciplinary effort, at this less pretentious level. The consideration of "points of articulation" among psychological, social, and cultural "systems" raises, at least obliquely, some interesting problems; the importance of making provisions, within approaches to psychological theory designed to cope with complex behavior, for variables adequate to the *empirical phenomena* designated by such terms as "the complementarity of expectations," "role-expectations," etc. is plausibly, if elliptically, argued. If the participants had kept more firmly in mind the "*Towards*" in "Towards a General Theory of Action," and had been less committed to the "*General*," and to the singular form of the autistic noun, "*Theory*"—the result might have been a useful preliminary mapping of common and differential problems of social science methodology.

Duke University. SIGMUND KOCH.

TINBERGEN, N. *The study of instinct.* London: Oxford Univer. Press, 1951. Pp. xii+228. $7.00.

It has sometimes happened in the history of science that a book has had influence greatly out of proportion to its apparent intrinsic importance. Paley's *Natural Theology*, permissible Sunday reading in a straight-laced time, influenced a whole generation of future English biologists by combining a wealth of natural history with its polemical "evidences of the existence and attributes of the Deity, collected from the appearances of Nature." Tinbergen's book may well serve a similar purpose for psychology. It comes as a partial answer to Beach's recent observations on the need for gathering behavioral data on a somewhat more representative number of species than American psychologists appear at present to be studying. *The Study of Instinct* should provide a means whereby the American student might become better acquainted with the increasing number of European studies of animal behavior and with the somewhat different approach to common problems which the zoologically-oriented ethologists are developing.

To those who have been reading the postwar journals *Behaviour* and the *Bulletin of Animal Behaviour*, or who have seen the *Fourth Symposium of the Society for Experimental Biology*, the general conceptual organization of Tinbergen's book will come as no surprise. Those who have confined their recent reading on animal behavior to the *Journal of Comparative and Physiological Psychology* and other American journals will find much that is different in concepts, in methods, and in factual description. Because the word "instinct" has acquired some of the connotations associated with "teleology," "reification," "purpose," and other such residues of the valiant battles of the Watsonian twenties, there is a real danger that the very people who can do most with Tinbergen's contributions will avoid them. It is hoped that this review will serve as some assurance that a book with "instinct" in the title can be at once worth while and exciting to the biologically-minded psychologist.

Tinbergen at the outset explicitly avoids the two traps which await every student of innate behavior. He refuses to accept the directiveness of Russell or the purposiveness of McDougall as explanation. While, as he says, the fact is undeniable, what the ethologist must study are the causal conditions of purposive behavior. He also rejects the subjective approach of Bierens de Haan that "an animal hunts because it is hungry," observing that this is "a guess about the possible nature of the animal's subjective state." What is needed is a "physiology of behaviour" that utilizes the objective methods of the physiologist but applies them to the problem of behavior rather than to the processes of organ systems as such. All of this is not, of course, startling, since any American physiological psychologist would probably say about the same things. However, Tinbergen next notes that the psychologist in America, although objective in methodology, has specialized on the higher types of behavior. Because of his interest in human conduct and his acceptance of man's evolutionary descent, he has become preoccupied with the *prehuman* behavior of mammals. His neglect of innate behavior, Tinbergen feels, "is due to the fact that it is not generally understood that learning and many other higher processes are secondary modifications of innate mechanisms, and that therefore a study of learning processes has to be preceded by a study of the innate foundations of behaviour."

Concerning methodology Tinbergen emphasizes "the importance of a complete inventory of the behaviour patterns of a species." Specialization on limited problems without knowing the behavior repertoire of the animal studied can lead to unwarranted generalizations, as can use of a few specialized techniques for all problems. "In view of the differences between any one species and another, the only thing that can be said for certain is that one should *not* use identical experimental techniques to compare two species, because they would almost certainly not be the same to *them*."

The major portion of the book is devoted to a survey of the roles of external stimulation and internal conditions responsible for innate behavior. The evidence is chosen from a wide variety of species and the insufficiency of merely determining the limits of sensitivity of receptors is emphasized, as follows:

> The animal's own world is not only dependent on what its sense organs can or cannot receive. Its sensory world is still more restricted; it is composed of sign stimuli, at least as long as we are dealing with innate responses. . . . The strict dependence of an innate reaction on a certain set of sign stimuli leads to the conclusion that there must be a special neuro-sensory mechanism that releases the reaction and is responsible for its selective susceptibility to such a very special combination of sign stimuli.

This process Tinbergen calls the Innate Releasing Mechanism. The extent to which this IRM is influenced by learning is looked upon as a matter for future research. Here Tinbergen might have profited by a more extensive search of the American literature on the "pre-human" behavior of such mammals as the rat, where good evidence in abundance may be found for a wide variety of acquired alterations of both stimulus pattern and IRM, particularly in such "innate" states as hunger and thirst.

His attempt at synthesis of the relevant data is along the general lines made familiar by Morgan's and Beach's concepts of internal mechanisms of motivation, coupled with a concept of hierarchical organization of the neural structures underlying innate behavior. As in all modern attempts to understand the physiology of need states, the endocrine system and hypothalamic nuclei are pressed into service. This material is presented with all modesty as a very tentative way of helping to "put our thoughts in order."

The role of learning in determining the precise sequence of these behaviors is again minimized, although other sections of the book present evidence to show that a great deal of individual modification of innate patterns does take place. Worthy of attention by American students of learning is Tinbergen's survey of what he calls the "innate disposition to learn." Many species show rapid and complex learning of some problems and almost complete stereotypy in other problems seemingly no more difficult. The mechanisms of naturalistic learning deserve more investigation.

It is difficult to make a critical evaluation of *The Study of Instinct* without retracing all of the polemical pathways of the 1920's. But it should be recognized that the investigation and the thinking of the European ethologists are far more rigorous than that of the field naturalists of the 19th century. Their concept of instinct, although similar to that of McDougall, is much better rooted in controlled observation. The final impression one gets from Tinbergen's book is that of the immense richness and incredible variety of behavioral phenomena which are there for the studying if we break out of our rat-monkey-man triangle. Tinbergen has no neat and tidy system and his problems sprawl over a range so wide as to appear to classify all behavior as "innate." But the facts he presents must be assimilated if the comparative method is to be something more than a misleading conventional symbol in American psychology.

University of Illinois. L. I. O'KELLY.

CAMERON, NORMAN, & MAGARET, ANN. *Behavior pathology*. New York: Houghton Mifflin, 1951. Pp. xvi+645. $5.00.

This book is a welcome departure from the great majority of texts in abnormal psychology in that the authors diligently attempt to present clinical material in terms of the interrelationships of developmental, learning, and biosocial factors. Whereas many authors give only lip service to the importance of these factors—their texts falling quickly into the Kraepelinean tradition despite prefatory statements to the contrary—Cameron and Magaret never seem to lose sight of the fact that these factors are always operative. The following statement by the authors indicates their approach as well as the scope of the book:

> We begin where birth begins, with the unmistakable needs of the neonate, and the succession of inescapable frustrations to which he is exposed because of his helplessness and the complexity of the social milieu. This points up the inevitability of conflict and anxiety; and it clears the way for our recognition that all persons must acquire techniques for overcoming anxiety or at least making it tolerable. The acquisition of such techniques and their modification are both examples of social learning; and this holds good, of course, for the acquisition and modification of pathological techniques, as well as normal ones, and for therapy as well as pathogenesis.

That this book is seriously concerned with general principles and constructs may be seen from the following topic headings: learning in a social situation, learning and temporal sequences of behavior, selective learning, language behavior, socially organized thinking, social role-taking.

It is perhaps unfair to have compared this book to other texts in abnormal psychology, most of which are adequate for a single course. This reviewer agrees with Carmichael's introductory statement: " . . . a possible value of this book would be lost if it were to become merely a useful professional guide to those who already have wide expert knowledge of man's normal and abnormal behavior. Actually the present volume is so clearly written and its scientific and yet eminently sensible point of view is so effectively presented that it is hard to believe that any educated man or woman who reads it can fail to acquire a new kind of understanding of why normal as well as abnormal people act as they do."

SEYMOUR B. SARASON.

Yale University.

SCHNEIDERS, ALEXANDER A. *Introductory psychology: The principles of human adjustment.* New York: Rinehart, 1951. Pp. xv+461. $4.00.

This book is comprehensive enough to give the student sufficient knowledge of the field, yet brief enough to be compressed into the standard three-hour one-semester course. The topics are organized around two fundamental psychological concepts, personality and adjustment. The first two chapters adequately introduce the student to psychology, first by discussion of what psychology is not, second by a systematic analysis of what psychology is. The latter includes the historical background and meaning of the term; representative viewpoints in modern psychology, including structuralism, functionalism and psycho-

Not all psychologists will want to treat "personality" as an independent entity, on equal footing with the biological heritage and environment. Some will prefer to regard it as the behavior outcome of a modifiable organism with a background of inherited characters in interaction with the environment.

The second part of the book is given over to the social psychology of age groups. However, the chapters given to infancy, childhood, adolescence, and the adult years and old age fail to reflect either the continuity of events which make up personality growth and development, or the effect upon such growth of the many social and social-psychological forces. True, the authors mention a few important developmental principles, such as "the organism develops as a unit," "maturation proceeds by coordination of two rather distinct processes, differentiation and integration," and "maturation is not uniform for all children." However, the specific researches which have given rise to these important concepts are almost completely neglected. For example, no material from developmental embryology appears. Shirley's important work is mentioned primarily to illustrate the fact that infants have little social interest in other infants.

In several places the concept of an emerging self is identified as a succession of social roles, one after the other. The authors acknowledge their intellectual debt to Cooley and Mead, who have undoubtedly afforded an important theoretical basis for a concept of the self. A philosophical position taken several decades ago, however important then in conceptualizing an undefined area, is hardly a substitute for the important empirical material which now undergirds a somewhat different concept of the emerging self. These authors make no mention of the important researches on the contribution of language to the development of the self concept.

In the discussion of adolescence, too, this reviewer is concerned that some of the most important papers on growth and adaptation during the adolescent years, such as those by Shuttleworth and by Jones, are missing from the references. What has been done is sound in a broad descriptive way. So much very important developmental material has, however, been omitted that the reader will gain no comprehension of the continuity of the process of growth.

Part III, much longer than Part II, deals with abnormal status, deviant roles, and personality. This reviewer believes the discussion of role status and the social psychology of minority status is one of the most comprehensive he has seen. Prejudice as a two-way process is excellently presented. The problem of group relationships is more objectively considered than in many current discussions which assume the majority to be "wrong" and the minority "right" when conflicts arise.

The final part has to do with social institutions and personality.

Here the school, religion, occupational status, the economic system, and the social psychology of mob behavior are considered. In a concluding chapter, the authors tackle the dimensions of the good society. Acknowledging the relativity of values, they define the good life in terms of the "golden mean of behavior" and venture to set up certain characteristics of a good society, although they recognize that good societies may take many forms.

This reviewer's criticisms thus far primarily reflect his expectation that the text would be developmental in its orientation. Viewing it as a contribution to social science, or more specifically to social psychology, he considers it a significant attempt to organize data concerning human behavior simultaneously from two traditionally different viewpoints. If the authors have been encyclopedic rather than critical in assembling and selecting their material, their attempt at synthesis should be viewed with some charity, for there is little intellectual precedent for their work.

One may rightly wonder where this book will be of greatest use. This reviewer believes that the typical undergraduate student in social psychology may find the book confusing. The intelligent undergraduate who has a broad background of reading in both sociology and psychology may find it stimulating. Graduate students with strong intellectual drives toward synthesizing concepts from many areas will also find it of interest. Psychological studies of concept formation affirm that adequate generalizations are made only in the presence of a rather large background of specific facts, data, ideas, and concepts. Hence a book which attempts to achieve the grand synthesis may properly be read after much specific course experience in sociology and psychology. Probably it should not be offered as a general text for the beginning student. The authors have made a good beginning toward bringing together and interpreting facts concerning human social behavior. The insights they develop reflect their own years of study; their product may not lead the sophomore who uses this as his social science text to a similar understanding.

DALE B. HARRIS.

University of Minnesota.

KUHLEN, RAYMOND G. *The psychology of adolescent development.* New York: Harper, 1952. Pp. xvii+675. $5.00.

In general, this textbook is solidly grounded in research and stresses practical applications. It includes critical sections, simply and briefly stated, with reference to most of the research quoted. By and large, theory is skirted. Psychoanalysis, for instance, is listed with three page entries in the subject index. The first two of these are brief and nonevaluative. No direct mention of psychoanalysis was found on the third page cited in the index. Freud is not listed in the author index.

The point of view is an eclectic, learning-organismic approach. The book presents broad (the range of references per chapter is from 24 to 94) and frequently unusual coverage of material.

The author points out repeatedly that data gathered on adolescents do not possess their full value in the absence of comparable data for adults. He emphasizes and documents consistently throughout the text cultural determination, multiple causation, and the functional interrelatedness of all aspects of development. He frankly includes certain ideological, thought provoking, "value including" sections (e. g., in the chapters dealing with American culture and personality and with adjustment to organized society).

There is a knowledgeable, practical section on the consequences of shifting to a democratic educational procedure from earlier, more formalized experiences. Unusual data on school variability, teachers' relations with students and their improvement, and a research and experience based section on grouping to reduce student variability help to make up one of the best chapters in the text (Chapter 10, "Educational development of adolescents and adjustment to school").

Effective use is made of some of the vivid material gathered by Baldwin, Kalhorn, and Breese on young children, to illustrate the range of home relationships. Brief case histories are used throughout. Sociometric data are discussed very practically, although the matrix presentation of such data is not covered. There is a brief discussion of autobiographical and other written materials as methods of studying individuals, although the emphasis on dangers and ethical considerations to be made in the use of such methods is not strong enough. Similarly, the section on the interview makes it look very simple; but interview ethics, and the need for intensive, supervised training in interview techniques are touched on only implicitly.

While the author demonstrates that he is aware of the operational meaning of "intelligence" he slips occasionally into phrases which carry excess meaning, such as "sheer underlying intellectual potential" (p. 87).

In the last chapter the treatment of test construction seems to the reviewer sketchy. It could lead to an uncritical acceptance of paper and pencil personality and adjustment tests, and an underestimation of the dangers of them in the hands of amateurs.

It seems to the reviewer, however, that the merits of the book clearly outweigh its defects. Factors of research orientation, clear but critical writing, and the author's obviously wide and "well-digested" practical experience make this a very suitable text for advanced undergraduates and beginning graduate students in education and human development.

BOYD MCCANDLESS.

State University of Iowa.

WATSON, ROBERT I. *The clinical method in psychology.* New York: Harper, 1951. Pp. xii+779. $5.00.

Watson's volume represents an excellent attempt to perform an impossible task. To undertake a text on clinical methods in psychology, pitched at both the "advanced undergraduate and graduate" levels, at the very outset requires courage. Along the way it calls for clinical and scholarly acumen, breadth plus intensity of knowledge, painstaking attention to tedious detail accompanied by ability to hold the reader's interest. These virtues are all reflected abundantly in this book. But in the end such an effort will inevitably be greeted by reviewers' dissatisfactions.

From the foregoing comments it should require no great skill in the clinical method to infer that the present reviewer feels somewhat guilty about offering several serious objections to an exceptionally sound piece of work. One immediate source of confusion concerns the audience which Watson is trying to reach. In his carefully elaborated "Apology to the Reader" he states that some previous acquaintance with diagnosis and therapy is taken for granted. For the undergraduate student such an assumption is likely to be ungrounded. For the graduate student some of the materials are too elementary, others seem to fit well into a first-year testing course, and still others would contribute to advanced study in the area. This uneveness of level probably means that each instructor will have to use the book selectively in terms of his own particular requirements.

The problem of balanced treatment of content also arises. Included are whole chapters on the Vineland Social Maturity Scale, diagnostic infant testing, and the Rosenzweig P-F Study, yet no discussion of such topics as interest tests and the Rorschach. The latter is excused, along with a number of other instruments, on the grounds of space-saving and also a contemplated separate volume, which of course leaves us with at least a temporary void. But apart from questioning the author's decisions on what to include and omit, the equal allocation of space to the Vineland Scale and the diagnostic interview, for example, appears dubious. On the positive side, it should be noted that the integration of diagnostic with therapeutic procedures has been accomplished in a highly effective manner.

Throughout the book there is a delicate weaving of research data into the pattern of clinical methodology. However, a notable exception to this approach is found in the absence of any research evidence bearing on the diagnostic interview. The negative findings of the VA Assessment Project, conducted by Kelly and Fiske, would provide a sober counterpart to the optimistic statement that the distinctive contribution of the interview is that it allows "the development of a total picture of the patient." Incidentally, the Assessment study, by far the most systematic investigation of clinical methods, does not even receive passing mention in the volume. Another overlooked study by Simos (*J. Pers.*, March 1950) would serve to temper the author's conclusions concerning the validity of the P-F test.

A text of this sort must draw its substance from the richness of the author's own knowledge and experience. Yet in a field as amorphous as clinical psychology, knowledge and experience cannot help but be contaminated by biases based on varying degrees of truth. Watson seems to have bent over backwards in trying to keep his own values out of the picture. But the danger is ever present that the subtle pseudopods of propaganda (most of which the reviewer is willing to endorse, by the way) may engulf the unsuspecting reader before he has a chance to think for himself. This dilemma is unavoidable. Occasionally more overt manifestations of the "projective hypothesis" creep into print, as in the author's defensive belaboring of the assertion that the clinician is really a scientist.

In summary, the book contains fine sections on diagnostic appraisal and psychotherapy. The latter occupies about a third of the total. Extensive bibliographies, including supplementary references from *Readings in the Clinical Method in Psychology*, edited by Watson, are also provided. There is some discrepancy between the front-flap quote that "liberal references to cases are supplied" and the author's own statement about the deliberate "relative paucity of case history material." In any event, the latter was not felt to be a deficiency. On the whole, it is easy to forecast that this book will be widely used, but somewhat more difficult to predict how often and for what reasons the clinical Sherlock will wish to consult Dr. Watson.

University of Michigan. GERALD S. BLUM.

PASCAL, GERALD, & SUTTELL, B. J. *The Bender-Gestalt Test: Quantification and validity for adults.* New York: Grune & Stratton, 1951. Pp. xiii+274. $6.60.

Presented here is a quantified system for scoring responses to the Bender-Gestalt Test. The authors found that in normally intelligent adults, free from brain damage, deviations in B-G performance are greatest in psychotics, less marked in neurotics, and smallest in normals. Data are presented which show the differences in performance of these three groups as well as the overlap.

For clinical practice and research, this book is enlightening insofar as it presents an attempt at quantification of the B-G, a detailed description of typical deviations, and numerous illustrations (including 96 complete records). The contributions of this work seem lessened, however, by the following objectionable features:

1. The B-G Test as applied by the present authors is, like most scales used

in clinical practice, an ordinal scale, with steps of unequal magnitude. This objection seems especially pertinent here because the items are differentially weighted.

2. Assuming that in adults without cortical damage, deviations in B-G performance are due to interpretative difficulties, rather than to perceptual or motor factors per se, the authors relate these difficulties to the subject's attitude toward the task. This attitude they relate, in turn, to the subject's attitude toward reality, which is believed to reflect his ego strength. The assumptions on which this new rationale is based are quite debatable.

3. In clinical practice, the B-G has been found to be sensitive to various cortical disturbances even of a mild nature. This new book contains little material that would aid the clinician in differentiating B-G deviations due to cortical deficit from those due to psychogenic disorders.

ELEANORE OCHS.

Bellevue Psychiatric Hospital, New Tork City.

RAPAPORT, DAVID. *Organization and pathology of thought.* New York: Columbia University Press, 1951. Pp. xviii+786. $10.00.

This is an interesting, informative, and quite scholarly source book of foreign literature dealing with the thought process and its pathology. The selections embrace psychology, psychiatry, and psychoanalysis, all brought together in an interesting fashion. Using the central theme of the motivation of thought processes as distinct from static analysis of thought mechanics, Rapaport has skillfully developed a symphony understandable to the sophisticate in all three fields.

Much of the material in this volume appears in English for the first time since uniqueness of resource was one of the criteria by which the author judged the suitability of inclusion. However, some notable exceptions which the dictates of central purpose made necessary add much to the continuity of text. For example some of Freud's work, though not translated for the first time, simply had to be included to round out the material.

The papers are grouped under five not necessarily mutually exclusive headings.

1. *Directed Thinking.* This section includes that material normally presented in classical psychology. The author points out that the difference between "Lewin on the one hand, and Ach and Buehler on the other, is as great as that between either of them and Claparède or Piaget, who are closer to clinical thinking than to experiment." The decision to omit some of the more important Gestalt contributions is regretted by the reviewer, although the physical dimensions of the project would have made their inclusion difficult.

2. *Symbolism.* In the reviewer's opinion this is the most valuable section in the book. More attention is given to this topic than to the others, which seems fortunate since modern trends in psychodynamics call for an understanding of symbol formation in the thought process. Some of the data reported by Silberer, Schroetter, Roffenstein, and Nachmansohn are both ingeniously

derived and surprisingly convincing to an academic psychologist. These experiments bear careful study and broader inclusion in our clinical teaching.

3. *Motivation of Thinking.* Here are included certain materials from Freud's *Collected Papers*, even though they had already been published in English. These inclusions were necessary for the development of a definitive volume. Also in this section is a very interesting paper by Heinz Hartmann on "Ego Psychology and the Problem of Adaptation."

4. *Fantasy Thinking.* Some papers of Bleuler, Varendonck, and Kris are included in this section.

5. *Pathology of Thinking.* Some of Schilder's work, as well as Bleuler's classical paper on "The Basic Symptoms of Schizophrenia" are the major contributions in this section. Bleuler's treatise is interesting to read, although much of it may be disputable in the face of modern evidence.

The author recognizes some of the volume's possible shortcomings. He has willfully theorized where the lack of objective evidence might have made some hesitant to speak. He has omitted the classical associationist, conditioned reflex, and conditioned response contributions because: "(a) my bias that these learning theories contribute little to a theory of thinking; (b) they are readily available, and my purpose was to gather those contributions which have received little attention; (c) I wished to avoid polemics in my comments." With the first of these reasons many will take exception. The classical research on conditioning seems like an oasis in a desert of psychoanalytic speculation to most psychologists. The third reason sounds strange in a theoretical document such as this, since most of the material might certainly be the subject of polemics. Nonetheless the volume is excellent, well written, scholarly, and certainly a mature piece of bibliographic research.

Marietta College. ROBERT P. FISCHER.

BRUNSWIK, EGON. *The conceptual framework of psychology.* Vol. I, No. 10 of International Encyclopedia of Unified Science. Univer. of Chicago Press, 1952. Pp. iv+102. $2.00.

Physically this is a slim volume, hardly a book, rather, a monograph of barely a hundred pages. Intellectually this is the equivalent of three books or, to put it conservatively, of one well-sized book and two monographs of about one hundred pages each. Such condensation, to whatever extent it can be achieved at all, is no mean feat and, like every achievement, has its price. Brunswik pays in both style and vocabulary. His style is overly parenthetical and allusive, his vocabulary formidably polysyllabic. Frequent use is made of small print and, worst of all, of an overelaborate terminology. All this has marred my pleasure. It may be that I am overreacting because I am so vividly reminded of a shared background which, for all its excellence, did not often excel in the art of saying things clearly and simply. But this will do for cavilling at what is, nevertheless, a distinguished piece of work by a distinguished author. So I turn to the three major themes, first to the two monographs and then to the book, that are here "encapsulated." (This is one of Brunswik's favorite words.)

There is, *first*, a reasoned statement of the logical positivists' analysis of psychology. Intermediate between epistemology and science proper in generality, this analysis belongs to what is now known as the philosophy of science. Since the *Encyclopedia* in which Brunswik's piece appears is an enterprise in this area, one might have expected him to devote all the space at his disposal to that theme. Had he followed this course, his statement of the position would probably have benefited. But I have no quarrel with it as it stands; at least, no quarrel that is not better confined to the epistemological seminar. The position expounded is also known as logical behaviorism or operationism. Brunswik's favorite name for it is objectivism, not unreasonably, since the term suggests the heart of the matter, namely, that psychology deals with overt behavior and nothing but overt behavior. However succinctly, he does give many of the explanations and qualifications this formula requires. And he also covers cursorily the other main tenets of the position by indicating the fallacies of the various methodological holisms and emergentisms, and by explaining how a "mechanistic" and "deterministic" psychology can deal adequately with purpose.

The *second* theme, or the second monograph the author might have written, is a piece of advocacy for a special line of research. Now Brunswik, who is a psychologist as well as a philosopher, is entitled to his opinions or hunches as to which kind of research is at the moment most promising. Yet, I cannot help feeling that the way in which he closely intertwines logical analysis with advocacy is most unfortunate. For, being merely a philosopher, I know only too well how hard it is to convince scientists that logical analysis properly conducted is genuinely neutral in that it does not and cannot furnish the grounds for such special pleading. As to Brunswik's preference, he wants all psychologists to look for laws that relate or, as he says, "focus on" two groups of variables, both "molar" and nonphysiological, the one "central," the other "distal." A need, for instance, is a central factor; test scores and perceptual "achievements" exemplify distal variables. Because of the nature and the multiplicity of the "mediating" mechanisms, laws of this kind are as a rule statistical or, as Brunswik puts it, "probabilistic." The early American school known as Functionalism was the first to investigate systematically relationships of this type. There is thus some historical justification for Brunswik's borrowing their name for his particular brand of psychology or, as I would rather say, his emphasis. For the rest, if I could permit myself the expression of a taste within science, I would tend to agree with this preference. So, by the way, do the students of Hull and Spence. Brunswik, rather amazingly, does not see that. This leads me to my last point.

The abstract of a book that is, *thirdly*, embedded in these hundred pages is a structural history of systematic psychology since Wundt and Brentano. The main task of a pragmatic history, such as Boring's, is the exposition of ideas in their causal sequence. Structural history is a comparative analysis of ideas in their logical interdependence. It is, no doubt, a discipline as admirable as it is difficult, requiring both considerable erudition and a philosophical turn of mind, a combination none too common. On the basis of the evidence presented, it is safe to say that few, if any, are better qualified than Brunswik to undertake this arduous task. But even structural history is a dangerous weapon in the hands of a man with a cause. In a pattern much too neat to be quite as thorough and exhaustive of the material as it looks, everything is made to point and converge toward the emergence of the new "functionalism," almost as the prophets of the Old Testament were ex post facto made into witnesses of the New. To write history in this vein is another less admirable trait of the background I share with the author. There is even an untranslatable German word for it (*Geschichtsklitterung*). I shall give two examples of the violence here perpetrated for the sake of a pattern. The one tries to eliminate a competitor; the other claims a spurious ancestor. The competitor, if this be the right word, is the behavior theory of the Hullians. This theory, we are told, since it "stays within the framework of associationistic tradition," can only give "token recognition" to stimulus patterning through "casually superimposed principles." As it happens, these principles are the very heart of the theory. Also, we are told that it can only interrelate its "intervening variables among one another and with peripheral-proximal boundary conditions," thus "paying with molecular encapsulation" for the progress it achieves. What, pray, is either molecular or peripheral or proximal about a rat's hunger and the piece of Purina it covets? The spurious ancestor of the new "functionalism" is the act psychology of Brentano. The one major historical root of Functionalism, particularly in its quasi-teleological aspects, is Darwinism. The only significant psychological issue at stake between the intentionalists and the Wundtians was the introspective indecomposability of what was later called imageless thought. True, Brentano was also a presentative realist; and in this epistemological sense his acts "intended" physical objects. But then, the task of the structural historian is to show how superficial the similarity is rather than to exploit it.

State University of Iowa. GUSTAV BERGMANN.

STONE, C. P., & TAYLOR, D. W. (Eds.) *Annual review of psychology.* Vol. 3. Stanford, Calif.: Annual Reviews, Inc., 1952. Pp. ix+462. $6.00.

Volume 3 of the recently initiated *Annual Review of Psychology* largely continues the pattern established by the first two volumes. It consists of 17 reviews of psychological literature covering the 1950–1951 period. A total of 462 pages represents a continuing quantitative gain over the 330 pages in Volume 1 and the 389 pages in Volume 2.

This volume, like the first two in the series, must certainly be recognized as an important and useful contribution. In the opinion of the present reviewer, however, it still falls somewhat short of the goals originally set by the Editorial Committee. Its shortcomings seem to lie principally in the failure of certain of its contributors to adopt a sufficiently evaluative and critical point of view with regard to the literature surveyed. The result is that some of the chapters tend to read like a kind of super-abstract, with strictly factual accounts of major publications strung together more or less like beads on a string. In all fairness it must be said that so extreme a categorization clearly applies only to a minority of the chapters and that another minority offers a sharp contrast by presenting exceptionally well-organized and critical evaluations of the literature.

An excellent example of effective evaluation of a major field is provided by Henry W. Nissen and Josephine Semmes in their review of comparative and physiological psychology. They include a formal section entitled "The Status of Comparative Psychology." Their chapter is marked also by its comprehensive coverage of certain of the recent important European works on comparative problems, especially that associated with the names of Lorenz and Tinbergen.

The other four chapters falling generally within the framework of experimental psychology offer a variety of approaches. Harry F. Harlow's capable review of the learning literature is well organized according to major research problems but is lacking in any formal introductory or overview material. Some of this kind of material, however, is interpolated throughout the chapter. The chapter on vision by Harry Helson is likewise well organized and has, in addition, more explicit introductory material. Both of these chapters are written in a lively tone and from a critical point of view. In these respects the review of hearing by Wendell R. Garner suffers by comparison. It is written in a matter-of-fact style with little attempt at evaluation of large trends. This is not to question its technical competence. The review of somesthesis and the chemical senses by G. R. Wendt is written in a somewhat similar style but does have a superior organization of topics. The wealth of detail in this chapter suggests that it is intended more for the specialist in the sensory field than for the nonspecialist.

A strong experimental emphasis is pleasantly evident in three borderline chapters—borderline in the sense of belonging not quite within the field of experimental nor quite within that of clinical or applied psychology. These are the reviews of child psychology by Vincent Nowlis and Helen H. Nowlis, of social psychology and group processes by M. Brewster Smith, and of motivation by O. H. Mowrer.

The stimulating chapter by the Nowlises is written from a strongly critical point of view and stresses the need for more systematic theoretical orientation in research on children. The chapter by Smith is introduced by a useful and interesting discussion of the present status of social psychology as a distinct discipline. The evaluative and critical tone is maintained throughout the remainder of the chapter. The editors are to be commended for the addition in this volume of the chapter on motivation. In this review Mowrer adopts a highly selective organization and so is able to cover only 56 references. The chapter is highlighted by relatively long but interesting sections on the historical development of motivation theory and the present status of the "punishment" problem, which together account for more than one-half of the review.

Four chapters may be classified as distinctly "clinical" in nature. The review of the literature on personality is by H. J. Eysenck of the University of London. It is one of two reviews by European psychologists, marking the initiation of the planned policy to include a wider representation than would presumably be provided by exclusive use of American reviewers. Eysenck's chapter is well prepared, with a highly personal touch, and is spiced with frequent critical comments. A similar personal organization is present in the review of abnormalities of behavior written by Joseph Zubin. This chapter is noteworthy for its clear-cut statement of the restricted purpose of the review and the basis of selection of the articles reviewed.

The two reviews of methodological literature in clinical psychology are less obviously personal in their organization and bias. Ann Magaret's otherwise excellent chapter on psychodiagnostics is marred only by her seeming reluctance to make critical comments pertaining to specific articles. This comprehensive review contains an exceptionally well-prepared overview of the field and thoughtful strictures on general methodological defects. The chapter on psychotherapy by Victor C. Raimy likewise contains a good overview and is written in a lively and witty manner.

One of the most stimulating chapters in the volume is that prepared by Lloyd G. Humphreys on individual differences, a subject matter that seems to belong both to clinical and applied psychology. Humphreys' explicit purpose is "to sketch an outline of a functional psychology of individual differences." He organizes the review around a number of selected research problems, indicates research needs as well as achievements, and does not hesitate to offer critical opinion of the articles reviewed. This chapter stands as a model of what can be achieved in an integration of the critical-evaluative and the pure abstracting functions.

Three reviews on distinctly applied subjects are offered. Clarence W. Brown and Edwin E. Ghiselli provide a top-notch chapter on industrial psychology, stressing research needs as well as critically evaluating achievements. Another competent review, surveying the literature on counseling, is contributed by William M. Gilbert. It should be noted that this chapter, carrying the title "Counseling: Therapy and Diagnosis," necessarily overlaps the two chapters on clinical methods (psychodiagnostics and psychotherapy). Perhaps the point of view is sufficiently different to justify the duplication, but it seems to this reviewer that a special editorial effort to delimit this area is indicated. The third applied review is that on educational psychology by John K. G. Elmgren, of the University of Gothenburg, Sweden. In certain respects this is the most unfortunate review in the volume. After a brief general discussion of the relationship between learning theory and educational problems, Elmgren launches into a technically competent but thoroughly humdrum and disconnected account of the various publications.

The single remaining chapter is a good, if somewhat condensed, review of statistical theory and research design by Quinn McNemar. The emphasis is on statistical *techniques*, as it should be, in spite of the fact that the title of the chapter refers to statistical theory. It is the opinion of the present reviewer that this kind of chapter stressing statistical and research techniques is today the most generally useful one to psychologists and should be expanded if possible in future editions.

Certain technical aspects of the 1952 volume deserve mention. One wonders why it should not be possible for the editors to devise a more logically planned arrangement for the order of appearance of the various chapters. No over-all organization is apparent in this respect, although short sequences of reviews of a similar type may be found. On the positive side, it is gratifying to note that, with but one exception, the bibliographies are now arranged alphabetically by author. Furthermore, the editors announce that on a trial basis in the next volume the reviewers will be given an opportunity to include if they will the full titles of journal articles. Reader opinion on this change is invited by the Editorial Committee, since it is stated that the extra space thus used will be in competition with that needed for exposition. (As a parenthetical suggestion in this connection, it may be ventured that considerable space could be saved if reviewers were discouraged from using the excessively long quotations—up to one-half page in length—that appear in several of the chapters.) It is certainly to be hoped that these corrections of irritating defects of the bibliographies in the early editions will become permanent features of the series.

It is also to be hoped that the critical emphasis on research methodology and research needs, especially evident in several of the chapters in the 1952 volume, will become increasingly evident in future volumes. The present reviewer sees this feature as the most important long-range contribution of the *Annual Review of Psychology* to the development of psychology, as the inevitable splitting up of the discipline into separated but interrelated subject-matter areas is intensified. It will certainly be a major contribution if this series aids in the cross fertilization within psychology of research problems and research techniques.

MELVIN H. MARX.

University of Missouri.

McGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* (2nd Ed., revised by A. L. Irion.) New York: Longmans, Green, 1952. Pp. xxii+596. $5.00.

In its original edition, *The Psychology of Human Learning* was a book which no serious student of the field could afford to be without. It constituted the best single bibliographic source available and contained McGeoch's always carefully considered evaluations of the status of most of the important problems in the area. Now, after ten years, Irion has undertaken the gigantic task of revising the work, trying simultaneously to include the most important developments contributed by some three thousand articles which have appeared in the last decade and to maintain the theoretical position (really the lack of it) espoused by the original author.

To a degree, these two objectives turn out to be incompatible. The decade from 1940 to 1950 was distinguished more by theoretical contributions than by additions to empirical subject matter in the field of human learning. Such developments have been handled in the revision in two ways: (*a*) two new introductory chapters (on theoretical matters and conditioning) have been included; (*b*) at several places, discussions of animal experimentation have been introduced, where they are considered enlightening. These sections, however, do not fit in well with the ultra-positivistic position supplied by McGeoch's point of view. The program of theory making involves the development of postulate systems which, by a process of *deductive* elaboration, lead to testable theorems which can be confirmed or disconfirmed by observations on behavior. Such a discussion is very difficult to gear to an essentially *inductive* point of view like that of McGeoch's.

These last comments should not lead to the conclusion that Irion's revision of McGeoch's book has turned it into a discussion of learning theory. They were made simply to show the kind of revision which ten years of development in the field of human learning has necessitated. Although they loom large in a discussion of *revisions*, they are less important in a description of the sort of book this is. It is still a scholarly, straightforward, factual summary of research in the field of human (chiefly verbal) learning. Considering the difficulties involved, Irion has done a remarkable job of maintaining the spirit of the original volume. And the reader's reaction to McGeoch and Irion will be about whatever it was to the original McGeoch.

A comparison of the literature cited in this book with that of the original edition provides us with an opportunity to assess the developments in the field of human learning that have occurred in the last ten years. Depending upon one's evaluation of the field in 1940, the result of such an analysis will be either comforting or distressing. For the book demonstrates that the mass of literature now available has not altered things markedly. A careful study reveals that only two new important contributors to the experimental literature (Irion and Underwood) have appeared in the last ten years. The authors leaned on most heavily in the original edition (Ebbinghaus, McGeoch, Melton, Thorndike) are still the important names in human learning. Similarly, the important topics are the same. The experimental problems treated in this book are very much like those discussed in the first edition. And the conclusions concerning matters of fact have changed little since 1940. Some changes in stress have occurred. The topics of reminiscence and the warm-up phenomena are given more attention in this book than in the previous one. This change has been chiefly at the expense of the topic of forgetting.

With one exception which will be pointed out below, the summary of the literature provided by McGeoch and Irion is excellent. And the question naturally arises as to the probable usefulness of a book of this sort. It is first of all *not* an undergraduate textbook. Too much background is taken for granted. For example, an understanding of operational definitions is assumed in the introductory chapter on theory. Furthermore, it shares with the original edition the characteristic of compressing a great deal of material into a fairly brief space. Undergraduates will find this volume hard going. Exactly the same features make the book a satisfactory (best available) text for graduate courses in human learning, although it has one chief drawback here too. This is the frequent reference to specific experiments (e.g., McGourty's master's thesis) as supporting some point without adequate description of the experiment. The student will find this mildly irritating or completely frustrating depending on his level of motivation.

As with the first edition of the book, McGeoch and Irion is a convenient source of summaries on a number of topics. The usefulness of the book in this way, however, will be somewhat restricted by the authors' policy of citing the *earliest* experiment where two or more equally good ones make the same point. Unless there is a very high correlation between recency of experimentation and excellence of the experimental procedure, this policy could lead to the systematic elimination of the most recent work in the field. No methodical attempt has been made to evaluate the seriousness of the result of this fact; but the reader should be aware that the most recent article cited by McGeoch and Irion is not necessarily the last word on the subject.

All in all, this revision lives up to the reputation of the original edition. The limitations, which have been pointed out above, do not weigh heavily against the fact that we now have a fairly up-to-date summary of an important field of psychology.

Duke University.

GREGORY A. KIMBLE.

KELLY, E. LOWELL, & FISKE, DONALD W. *The prediction of performance in clinical psychology*. Univer. of Michigan Press, 1951. Pp. xv+311. $5.00.

> During the Spring of 1946, the Veterans Administration completed its plans for training large numbers of clinical psychologists, . . . the Veterans Administration decided also to institute a program of research designed to make possible a more effective selection of trainees than might otherwise occur. . . . Early in the summer of 1946, the University of Michigan was approached as to its willingness to accept the responsibility for the direction and coordination of such a research program. . . .
>
> . . . The over-all design of the project was as follows: in 1947 and 1948, several hundred college graduates seeking admission to or just entering the four-year VA training program in clinical psychology in some 40 universities were evaluated by a wide variety of techniques, and predictions were made concerning their probable success in training and their future professional competence. . . .

The Prediction of Performance in Clinical Psychology is an account of a necessarily elaborate study which is remarkable both for the candor of its presentation and the breadth of its content. All of the various phases, which required a total of five years, are described in useful detail. In essence clinical psychologists have employed the most promising of their devices in an attempt to predict the responses of students to training in clinical psychology. Although the most hoped for goals in this study were not realized, the scope of this report makes it an important chapter in our literature of applied psychology.

This study has succeeded in providing us with additional reassurance that we can make useful predictions of a student's achievements with respect to the more formal academic aspects of graduate training. The results for the primarily clinical criteria do not provide a means for useful prediction of the development of clinical students as clinicians.

Considering the intangible nature of clinical proficiency and the general lack of satisfactory evidence concerning the efficacy of clinical procedures per se, this failure is not surprising. Accordingly, the following comments are neither criticisms nor suggested improvements; they merely provide a context for a brief discussion of the general nature of some features of this interesting work.

The author's report describes two major emphases. One has to do with the development of measurements and indices which are of predictive value; the other has to do with the development of measurements and indices which are criteria for the development or attainments of student clinical psychologists. From the standpoint of the predictive variables, the undertaking was unique in both its extensity and intensity. Tests of mental ability, of various achievements, and of vocational interests, as well as personality attributes, and judgments and ratings based on subjective tests, interview situations, and sociometric procedures were all employed. It is probable that the various clinical psychologists who participated in this undertaking (the list of participants looks like a "Who's Who in Clinical Psychology") applied their clinical skills and insights unsparingly to this task.

The emphasis on the criteria is frank and conscientious. As a matter of fact the sections which have to do with the development of criteria could serve as a reference for the current definition of clinical psychology. It seems to this reviewer, however, that the manner in which the criterion problem was handled was determined by one major point of view. Specifically, it appears that the investigators felt either that they knew or ought to know the desirable characteristics of a clinical psychologist. (This quite natural bias is explicitly acknowledged by the authors.) This bias led to the development of criteria which emphasized "goodness" among the developing clinical psychologists rather than "differences" between them. This necessarily meant that attempts to make useful predictions for the developing students were limited to a restricted number of currently acceptable aspects of "goodness." As a consequence no comprehensive prediction of "differences" is attempted.

Perhaps the major clinical criteria were ratings of diagnostic ability, therapeutic ability, and desirability as an employee. Although it appears that such ratings may be made with fair consistency, one is inclined to wonder how much the meaning of these ratings could vary from school to school and whether or not there are internal consistency analyses to indicate that all of the items on the rating scales belong on the same continuum. It is possible that the items on the rating scales are very well placed from the standpoint of general desirability, i.e., "goodness." It is also possible that from a behavioral standpoint they do not belong on the same scale.

It appears that individuals from different schools were combined for the purpose of statistical analysis. Nevertheless, it is known that schools differ in the manner in which they select students and the manner in which they train students. This suggests a possibility that for some variables the magnitude or the direction of correlation may vary somewhat from school to school. It also seems possible that a respective correlation for a sample of schools and for a sample of individuals within a school might be different. Although the writers were obviously and most properly addressing themselves to trends which obtain for all VA clinical trainees, it is quite possible that the essential nature of such trends could be found by summarizing the possible trends found separately in individual schools (or in groups of similar schools) and that such a summary could present a more optimistic picture than one revealed for the total sample of all schools. As a matter of fact, it may be difficult for some readers to decide whether they should attempt to generalize to schools or to individuals. An attempt was made to eliminate school differences from some of the variables and it is possible that this reduced the heterogeneity of the measures for the total group, thus reducing the probability of a high correlation.

The authors have been conservative in all of their claims and have not made predictions based on weighted combinations of various measures; presumably, they sought to avoid a spuriously high indication of relationship. It seems possible, however, that some weighted combination of the data could be profitable. It is hoped that for a restricted portion of the sample multiple predictions would be prepared and then cross-validated on some similar but independent portion of the sample.

The present review does not presume to provide a detailed commentary on the methods and rationale of the study, nor is it intended to propose alternative approaches. Nevertheless, it is quite possible that many readers would like to know why alternative analytical procedures were not employed or why a more diversified attempt at the criterion problem was not attempted. Accordingly, the negative implications of the present study should not be overemphasized. Perhaps meager results should never be taken as an indication that positive results are not possible. Certainly the present study should not be taken as a basis for claiming that important aspects of clinical proficiency cannot be anticipated from the use of these or other predictive measures. On the contrary, many aspects of the present study provide a useful and encouraging exploration.

Yale University. J. R. WITTENBORN.

BOOKS AND MATERIALS RECEIVED

ANDERSON, IRVING H., & DEARBORN, WALTER F. *The psychology of teaching reading.* New York: Ronald, 1952. Pp. x+382. $4.75.

ARTHUR, GRACE. *The Arthur adaptation of the Leiter International Performance Scale.* Washington, D.C.: Psychological Service Center Press, 1952. Pp. viii+73.

AUSUBEL, David P. *Ego development and the personality disorders.* New York: Grune & Stratton, 1952. Pp. xi+564. $10.00.

BECK, S. J. *Rorschach's test.* Vol. III. *Advances in interpretation.* New York: Grune & Stratton, 1952. Pp. viii+301. $5.50.

BENEDEK, THERESE. *Studies in psychosomatic medicine—psychosexual functions in women.* New York: Ronald, 1952. Pp. x+435. $10.00.

BERRIEN, F. K. *Practical psychology.* (Rev. Ed.) New York: Macmillan, 1952. Pp. ix+640. $5.00.

BRAIN, SIR RUSSELL. *The contribution of medicine to our idea of the mind.* New York: Cambridge Univer. Press, 1952. Pp. 29. $.50.

CHESSER, EUSTACE. *Cruelty to children.* New York: Philosophical Library, 1952. Pp. 159. $3.75.

CLEUGH, JAMES. *The Marquis and the Chevalier.* Boston: Little, Brown, 1952. Pp. 295. $4.50.

CUSSLER, M., & DEGIVE, M. *'Twixt the cup and the lip.* New York: Twayne Publishers, 1952. Pp. 262. $3.95.

EDUCATIONAL RECORDS BUREAU. *1952 achievement testing program in independent schools and supplementary studies.* Bulletin No. 59. New York: Educational Records Bureau, 1952. Pp. xiii+105.

EISLER, R. *Man into wolf.* New York: Philosophical Library, 1952. Pp. 286. $6.00.

FARIS, ROBERT E. L. *Social psychology.* New York: Ronald, 1952. Pp. vii+420. $5.00.

GORLOW, L., HOCH, E. L., & TELSCHOW, E. *The nature of nondirective group psychotherapy.* New York: Bureau of Publications, Teachers Coll., Columbia Univer., 1952. Pp. viii+143. $3.25.

GREENE, EDWARD B. *Measurements of human behavior.* (Rev. Ed.) New York: Odessey, 1952. Pp. xxiv+790. $4.75.

GUILFORD, J. P. *General psychology.* (2nd Ed.) New York: Van Nostrand, 1952. Pp. xii+587. $5.00.

HYNDMAN, OLAN R. *The origin of life and the evolution of living things.* New York: Philosophical Library, 1952. Pp. xxi+648. $8.75.

JAQUES, ELLIOTT. *The changing culture of a factory.* New York: Dryden, 1952. Pp. xxi+341. $4.25.

INDEX OF SUBJECTS

APA Publication Manual, 389
Analysis of variance, a qualification in the use of,
  Diamond's comment on Webb and Lemmon, 151
  Patterson's note on Webb and Lemmon, 148
  Webb and Lemmon's sequel to the notes of Patterson and Diamond, 155
Analysis of variance designs for learning studies, greco-latin, 521

Behavior (social) of vertebrates: review of the literature (1939–1950), 598
Blacky Pictures, the validity of, 238

Calculating correlation coefficients, the effectiveness of short methods for, 342
Christie's "Experimental naïveté and experiential naïveté," note on, 38
Comparative psychology, conceptual trends in, 559
Computational short cut in factor analysis, 144
Conceptual trends in comparative psychology, 559
Correlation coefficients, the effectiveness of short methods for calculating, 342
Criteria (experimentally controlled), evaluation of, 333

Educational and psychological testing, item analysis in relation to, 97
Evaluation of experimentally controlled criteria, 333
Experiential naïveté, experimental naïveté and, note on, 38
"Experimental naïveté and experiential naïveté," note on, 38
Experimentally controlled criteria, evaluation of, 333

Factor analysis, computational short cut in, 144
Factor-analytic research designs—their interrelations and derivatives, 499
Factor analyze, when not to, 26
Functional relations obtained from group data, note on, 263

"Genetics of schizophrenia"
  Hurst's further rejoinder to Pastore, 544
  Pastore's rejoinder to Hurst, 542
Greco-latin analysis of variance designs for learning studies, 521
Group data, note on functional relations obtained from, 263
Group discussion technique (leaderless), the reliability of, 540
Guertin, Rabin and, "Research with the Wechsler-Bellevue test," correction, 71

Homogeneity-heterogeneity in the use of matching method in personality studies, note on the problem of, 41
Hypothesis testing, one-sided vs. two-sided alternatives, 43

Item analysis in relation to educational and psychological testing, 97

Leaderless group discussion technique, the reliability of, 540
Learning studies, some greco-latin analysis of variance designs for, 521
Latent learning
  Blodgett and Haney types of latent learning experiment: reply to Thistlethwaite, 52
  Comment on Thistlethwaite's perception of, 47
  Thistlethwaite's reply to Kendler and Maltzman, 61

Matching method in personality studies, problem of homogeneity-heterogeneity in the use of, 41
Meaning, nature and measurement of, 197
Measure of relation determined by both

mean difference and profile information, 251
Measurement of meaning, the nature and, 197

Naïveté, experimental and experiential, note on, 38
Nature and measurement of meaning, 197
Non-parametric statistics for psychological research, 122

Personality studies, problem of homogeneity-heterogeneity in the use of matching method in, 41
Personality test responses, influence of social class on, 318
Profile information and mean difference, a measure of relation determined by, 251
Profile similarity, note on, 538
Prognosis (psychopathological), psychological tests in, 451
Psychological research, non-parametric statistics for, 122
Psychological stress, effects of upon performance, 293
Psychological testing, item analysis in relation to, 97
Psychological tests in psychopathological prognosis, 451
Psychopathological prognosis, psychological tests in, 451
Publication Manual of the APA, 389

Q technique, some observations on, 483

Rabin and Guertin's "Research with the Wechsler-Bellevue test," correction, 71
Ratings, note on the combination of, on the basis of reliability, 270
Reliability, note on the combination of ratings on the basis of, 270
Reliability of the leaderless group discussion technique, 540
Reliable and questionable significance in a series of statistical tests, 339
Research designs, factor analytic, 499
Review of the literature on social behavior of vertebrates (1939–1950), 598

Schizophrenia, genetics of
  Hurst's further rejoinder to Pastore, 544
  Pastore's rejoinder to Hurst, 542
Short methods for calculating correlation coefficients, the effectiveness of, 342
Social behavior of vertebrates: a review of the literature (1939–1950), 598
Social class, influence of, on personality test responses, 318
Stress (psychological), effects of upon performance, 293
Statistical tests, reliable and questionable significance in a series of, 339
Statistics (non-parametric), for psychological research, 122

Test (personality) responses, influence of social class on, 318
Tests of hypotheses; one-sided vs. two-sided alternatives, 43
Thematic Apperception Test: interpretive assumptions and related empirical evidence, 1

Validity of the Blacky Pictures, 238
Vertebrates, review of the literature on the social behavior of (1939–1950), 598

Wechsler-Bellevue test, research with (1945–1950), correction, 71

INDEX OF AUTHORS

GENERAL REVIEWS, SHORT ARTICLES, SPECIAL REVIEWS, NOTES

Andree, R. V., 144
Archer, E. James, 521
Auld, Frank, Jr., 318
Bernstein, Lewis, 38
Blum, Gerald S., 238
Brožek, Josef, 339
Cattell, Raymond B., 499
Davis, Frederick B., 97
Deese, James, 293
Diamond, Solomon, 151
Flanagan, John C., 342
Grings, William W., 333
Guertin, W., 71
Guilford, J. P., 26
Hunt, Howard F., 238
Hurst, Lewis A., 544
Jones, Lyle V., 43
Kendler, Howard H., 47
Lawshe, C. H., 270
Lazarus, Richard S., 293
Lemmon, Vernon, 155
Lindzey, Gardner, 1
Maltzman, Irving, 52
Moses, Lincoln E., 122
Nagle, Bryant F., 270
Osgood, Charles E., 197, 251
Osler, Sonia F., 293
Pastore, Nicholas, 542
Patterson, C. H., 148
Rabin, A., 71
Ross, Sherman, 598
Schneirla, T. C., 559
Secord, Paul F., 41
Semeonoff, Boris, 540
Sidman, Murray, 263
Smith, Wendell I., 598
Stephenson, William, 483
Suci, George J., 251
Thistlethwaite, Donald, 61
Tiede, Kenneth, 339
Webb, Wilse B., 155
Webster, Harold, 538
Windle, Charles, 451

BOOKS REVIEWED

Arnold, Wilhelm, 88
Barker, Roger G., 92
Bartley, S. Howard, 80
Beach, F. A., 75
Berrien, F. K., 377
Blake, R. R., 72
Bohm, Ewald, 287
Bowles, J. W., Jr., 86
Brunswik, Egon, 654
Bugelski, B. R., 82
Cameron, Norman, 644
Cattell, Raymond B., 89, 90
Cook, Stuart W., 549
Culbertson, James T., 384
Davis, A., 370
Deutsch, Morton, 549
Dewey, Richard, 647
Dollard, John, 183
Eells, K., 370
Ellis, Robert S., 379
English, Horace B., 380
Fiske, Donald W., 662
Ford, C. S., 75
Garrett, Henry E., 85
Gurvitz, Milton S., 372
Havighurst, R. J., 370
Helson, H., 349
Herrick, V. E., 370
Hofstätter, Peter R., 646
Horrocks, John E., 381
Hull, Clark L., 628
Humber, W. J., 647
Humphrey, George, 547
Irion, A. L., 660
Jahoda, Marie, 549
Janis, Irving L., 372
Jones, L. Wynn, 193
Katona, George, 276
Keller, Fred S., 189
Kelly, E. Lowell, 662

Kempthorne, O., 548
Kuhlen, Raymond G., 649
Lindquist, E. F., 194
Luneburg, Rudolf K., 363
McClelland, David C., 367
McGeoch, J. A., 660
Magaret, Ann, 644
Marx, Melvin H., 355
Miller, George A., 361
Miller, J. G., 185
Miller, Neal E., 183
Moreno, J. L., 283
Morgan, Clifford T., 78
Munn, Norman L., 357
Murphy, Gardner, 83
Parsons, Talcott, 636
Pascal, Gerald, 652
Pronko, N. H., 86
Queener, E. L., 282
Ramsey, G. V., 72
Rapaport, David, 653
Rogers, Carl R., 374
Rohrer, John H., 280
Schneiders, Alexander A., 645
Schoenfeld, William N., 189
Sherif, M., 280
Shils, Edward A., 636
Spearman, C., 193
Stellar, Eliot, 78
Stevens, S. S., 156
Stone, Calvin P., 359, 657
Suttell, B. J., 652
Symonds, P. M., 382
Symposia for the Society for Experimental Biology, 385
Taylor, D. W., 657
Thompson, Laura, 187
Tilton, J. W., 551
Tinbergen, N., 642
Tolman, Edward C., 274
Tyler, R., 370
Volkart, Edmund H., 284
Watson, Robert I., 651
Welford, A. T., 376
Wright, Herbert F., 92
Yacorzynski, G. K., 92

FILMS REVIEWED

Bennett, A. E., 557
Hayes, C., 555
Hayes, K. J., 555
Hunt, Gardner L., 554
McKeever, L. G., 557
Miller, Neal E., 554
Pronko, N. H., 553
Snyder, F. W., 553
Stone, L. J., 556
Verplanck, W. S., 557

BOOK AND FILM REVIEWERS

Amatora, S. M., 645
Anderson, John E., 162
Baldwin, Alfred L., 90
Barmack, Joseph E., 78
Beach, Frank A., 385
Bennett, Edward M., 555
Bergmann, Gustav, 654
Bersh, P. J., 628
Blair, Glenn M., 381
Blum, Gerald S., 651
Brown, Judson S., 167, 189
Burke, C. J., 548
Butler, John M., 183
Ericksen, S. C., 85
Estes, William K., 355
Finger, Frank W., 83
Fischer, Robert P., 653
Garner, W. R., 176, 179
Gilchrist, J. C., 549
Glixman, Alfred F., 185
Graham, C. H., 363
Grant, David A., 158, 169
Guetzkow, Harold, 280
Harding, John, 283
Harris, Dale B., 647
Hayes, Samuel P., Jr., 276
Heathers, Glen, 187
Hunt, William A., 372
Jenness, Arthur, 382
Johannsen, Dorothea E., 553
Johnson, Donald M., 82
Kappauf, William E., 180

Katz, Elias, 556, 557
Kendler, Tracy S., 554
Kennedy, John L., 160
Kimble, Gregory A., 660
Koch, Sigmund, 349, 636
Lanier, Lyle H., 156
Leeper, Robert, 367
Lehner, George F. J., 86, 646
Leuba, Clarence, 282
McCandless, Boyd, 649
McClelland, David C., 72
Macfarlane, Jean Walker, 89
McGuigan, F. J., 384
MacLeod, Robert B., 274
McNemar, Quinn, 370
Marx, Melvin H., 657
Miles, Dwight W., 374
Miller, George A., 547
Miller, James G., 183
Muenzinger, Karl F., 88
Ochs, Eleanore, 652
O'Kelly, L. I., 642
Osgood, C. E., 361
Piotrowski, Zygmunt, 287
Pfaffmann, Carl, 80
Ratoosh, P., 363
Richards, T. W., 380
Riggs, Lorrin A., 174
Sarason, Seymour B. 92, 644
Sargent, S. Stansfeld, 377
Saunders, D. R., 193
Schoenfeld, W. N., 628
Shock, Nathan W., 376
Smith, M. Brewster, 372
Snygg, Donald, 379
Stellar, Eliot, 75, 165, 179
Stendler, Celia Burns, 92
Stroud, J. B., 551
Travers, Robert M. W., 194
Ullman, Albert D., 557
Williams, Stanley B., 357
Wittenborn, J. R., 662
Young, Kimball, 284
Zeaman, David, 359